Evaluating Near Infrared Hyperspectral Imaging as a Complimentary Rapid Screening Tool for Food Microbiology

by

Celesté Nadine Mapling

Thesis presented in partial fulfilment of the requirements for the degree of

Master of Science in Food Science

at

Stellenbosch University

Department of Food Science, Faculty of AgriSciences

Supervisor: Dr Paul J Williams

March 2020

DECLARATION

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: March 2020

ABSTRACT

Near infrared hyperspectral imaging (NIR-HSI) was evaluated as a complimentary, rapid screening tool for microbiology. Six bacterial isolates were used including *Bacillus cereus* (ATCC 13061), *Escherichia coli* (ATCC 25922), *Salmonella enteritidis* (ATCC 13076), *Staphylococcus aureus* (ATCC 29213 & 25923) and *Staphylococcus epidermidis* (ATCC 12228). The bacteria were streaked out onto nutrient agar (NA) and tryptic soy agar (TSA) to evaluate the effect of different growth media on the NIR spectra. For the second objective both a streak and spread plate method was used on only NA to evaluate the effect of the plating methods on the spectra. The bacteria were classified based on their Gram-stain classification (Gram-positive or Gram-negative) and based on pathogenicity (pathogenic and non-pathogenic). Hyperspectral image analysis was conducted in the 950-2500 nm range. The images were collected using the Hyspex SWIR-384 push-broom imaging system. Two pre-processing methods were applied including standard normal variate (SNV) and Savitzky-Golay (2^{nd} derivative, 3^{rd} polynomial,) to assist with image cleaning. The images were mosaicked into groups based on the overall objective of the analysis. Principal component analysis (PCA) models were calculated for initial data exploration and reduction of data dimensionality. Partial least squares discriminant analysis (PLS-DA) models were then calculated to classify the individual bacteria based on Gram-stain reaction and pathogenicity. The PLS-DA model of the bacteria streaked out on TSA produced the best overall classification accuracy (100%). The bacteria on NA obtained a classification accuracy of 99.55% and the Gram-stain classification and pathogenicity models produced classification accuracies of 98.29% and 98.72%, respectively. For the second objective the bacteria were classified based on Gram-stain reaction. To further optimise the model, three different pre-processing combinations were employed. The PLS-DA model containing only Gram-negative bacteria achieved a 100% classification accuracy, followed by 99% for Gram-positive bacteria and 89% for the model combined. This proves that hyperspectral imaging can successfully be implemented as a rapid screening tool to accurately detect and identify foodborne pathogenic bacteria.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank and acknowledge our heavenly father for blessing me with the strength, knowledge and ability to learn and undertake this research study. Without His endless grace and faithfulness, I would not have been able to persevere and reach this lifelong goal.

To my supervisor, Dr P J Williams, who believed in me and always pushed me to believe in myself and be confident in my abilities, I am extremely grateful. This project was not an easy task at all considering the many challenges we faced but knowing that I had your support throughout, definitely lightened the load. Thank you for all the motivational pep-talks and knowledge you have shared over the past two years.

To all the staff at the Department of Food Science who have extended a helping hand throughout this journey, I am immensely grateful. I would also like to express my sincerest gratitude to the Head of the department, Prof G O Sigge, for the financial support I have received this year, I could not have successfully completed my studies without it.

A very special thank you to Dr J Colling, to whom no task was too big to help with. Thank you for your kind heartedness and willingness to help, I am so grateful to have had your support this year.

The National Research Foundation (NRF) [Free-standing Scholarships (2018) & Scarce Skills Master's Scholarship (2019)] is hereby acknowledged for financial support (any opinion, findings and conclusions or recommendations expressed in this material are those of the author and therefore the NRF does not accept any liability in regard thereto).

I would also like to thank my friends and family, who have supported me throughout my academic career, for always pushing me and motivating me when it was hard to motivate myself. Thank you to Angel, for the endless laughter (and tears), for the late nights and early mornings, your support is highly appreciated.

Lastly, to my mother, Muriel Sophia Mapling, who raised me singlehandedly and always pushed me to be and give my absolute best. Mom, without your prayers, love and support I would not have been able to complete my studies. You truly are my rock and I dedicate the hard work of past two years and this degree to you. To my sister, Megan Nealicia Mapling, thank you for your undying support over the years. Whether it be financially, emotionally or by travelling from the Middle East to physically be here to celebrate my achievements, I truly appreciate everything you've done for me. My number one supporters, EK LIEF JULLE ONEINDIG BAIE!

LIST OF FIGURES

LIST OF TABLES

CONTENTS

LIST OF ABBREVIATIONS

2-D	Two Dimensional Three Dimensional
3-D	Three Dimensional Three Dimensional
ANN	Artificial Neural Networks
AOTF	Acousto-Optic Tunable Filters
BHI	Brain heart infusion
CCD	Charge Coupled Device
CMOS	Complementary Metal Oxide Semiconductors
ELISA	Enzyme-Linked Immunosorbent Assay
EMCCD	Electron Multiplying Charge Coupled Device
EMSC	Extended multiplicative scatter correction
FOV	Field of view
FT-IR	Fourier Transform Infrared
FT-NIR	Fourier Transform Near Infrared
HCA	Hierarchical Clustering Analysis
HMI	Hyperspectral microscope imaging
HSI	Hyperspectral Imaging
InGaAs	Indium Gallium Arsenide
kNN	k-nearest neighbour
LB	Luria-Bertani
LDA	Linear Discriminant Analysis
LED	Light Emitting Diodes
LOD	Limit of detection
LV	Latent Variable
MDA	Multivariate Data Analysis
MHA	Mueller Hinton Agar
MIR	Mid-infrared
ML	Maximum likelihood
MLR	Multiple Linear Regression
MRS	De Man, Rogosa and Sharpe
MSA	Mannitol salt agar

MSC	Multiplicative Scatter Correction
NA	Nutrient agar
NIR	Near Infrared
NW	Norris-Williams
PC	Principal Component
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PDA	Potato Dextrose Agar
PLS	Partial Least Squares
PLS-DA	Partial Least Squares Discriminant Analysis
PLSR	Partial Least Squared Regression
PPC	Psychrotrophic plate count
R^2	Coefficient of Determination
ROI	Region of Interest
RS	Raman Spectroscopy
SG	Savitzky-Golay
SIMCA	Soft Independent Modelling of Class Analogy
SNR	Signal to noise ratio
SNV	Standard Normal Variate
STEC	Shiga toxin-producing *Escherichia coli*
SWIR	Short Wave Infrared
TH	Tungsten-Halogen
TSA	Tryptic soy agar
TVC	Total viable count
UV	Ultraviolet
Vis-NIR	Visible-Near Infrared
VIP	Variable importance in projection
VRBA	Violet red bile agar
WHO	World Health Organization
ZnSe ATR	Zinc selenide attenuated total reflectance

CHAPTER 1

Introduction

Food safety is regarded as one of the most prominent technological and regulatory challenges globally. Various infectious agents including bacteria, fungi, viruses, and parasites are responsible for causing over 200 known diseases (Oliver *et al.,* 2005; Bhardwaj *et al*., 2017). Furthermore, food safety affects at least four out of the eight established Millennium Development Goals (MDG) set out by the World Health Organization (WHO). Regulation and control systems developed specifically for the mitigation of food safety issues are on the radar as these goals are evidently not being met. The financial burden placed on governmental control systems are further exacerbated by challenges that include population expansion, mass product distribution and ultimately product recalls among others (Aluwong & Bello, 2015).

Foodborne diseases (FBD) are an important cause of morbidity and mortality worldwide (WHO, 2015). Among the most common foodborne pathogens responsible for foodborne outbreaks are *Listeria monocytogenes*, *Escherichia coli* O157:H7 and Shiga toxin-producing *E. coli* (STEC) *Staphylococcus aureus*, *Salmonella enterica*, *Bacillus cereus*, *Vibrio* spp., *Campylobacter jejuni* and *Clostridium perfringens* (Law *et al*., 2015). Moreover, an immense amount of pressure is being placed on the food industry to develop improved early detection methods to ultimately eradicate foodborne illnesses.

Over the years the food industry has relied heavily on culture-based microbiological and biochemical methods to detect and identify food related microorganisms. However, although traditional culture-based methods are reliable (Bhardwaj *et al*., 2017), they are limited in their ability to accurately detect organisms to a species level (Blackburn, 2006). These methods include laborious plating techniques and enumeration procedures by specialized trained staff and causes unavoidable delays. This also results in large volumes of products being distributed before being tested for possible contamination (Kammies *et al*., 2016). Exploring rapid methods that could enhance current conventional methods when used in conjunction, is becoming of vital importance for the food industry (Wang *et al.*, 2018a).

A vast amount of rapid and automated methods has been developed for the identification of bacteria which include serological and molecular approaches. This include immunological-based (e.g. enzyme-linked immunosorbent assay (ELISA), nucleic acid-based (e.g. PCR)) and biosensor-based (e.g. optical biosensors) methods among others (Law *et al*., 2015). These methods have been extensively reviewed (Zhao *et al*., 2014; Law *et al*., 2015; Wang *et al*., 2016). The requirements for rapid methods are that they are labour-saving, generally more sensitive and time-efficient, with an added reliability to enhance the use of current conventional methods (Law *et al*., 2015).

Spectroscopic methods have also been developed for detection of foodborne pathogens. The near infrared region (NIR) region (700- 2500 nm), of the electromagnetic spectrum, has gained much interest and has been used for a variety of applications. Sir William Herschel first discovered light in the NIR-region in the

1800's, however the earliest applications were only reported in the 1950's. It was not until the 1970's that it was first used in the agricultural industry (Cen & He, 2007). The advantages of the technique include gaining information on the chemical and physical state of a sample from a single spectrum, which ultimately permits time efficient analysis. However, NIR spectroscopy is based on measuring bond vibrations within an organic molecule. It is regarded as a qualitative tool that provides empirical data and therefore involves intricate calibration procedures and chemometrics (McClure, 2003). The technique relies on mean spectra, which poses additional challenges when analysing complex heterogeneous samples.

One solution to this challenge is NIR-hyperspectral imaging (NIR-HSI). This non-destructive technique combines spectroscopy and chemical imaging and is based on the unique ability of microorganisms that are able to provide distinct NIR profiles (Wang, 2018). These profiles can be used to 'map' distinct biochemical patterns or fingerprints used for classification (Baldauf, *et al*., 2006). Hyperspectral images are acquired using three general configurations namely push-broom, whiskbroom and stare-down. This produces three-dimensional images, known as hypercubes, which contain both spectral and spatial information about a target sample. Operating the instrument provides ease of use and images are typically acquired in less than one minute. Combining these capabilities with adequate data processing methods is regarded a value-added ability to the rapid method development world (Amigo *et al*., 2010).

The general steps in analysing hyperspectral images include spectral pre-processing techniques, image compression and analysis. Multivariate data analysis (MDA) techniques, such as principal component analysis (PCA), are used to reduce the dimensionality of various datasets and is usually referred to as the workhorse of MDA (Geladi *et al.*, 1989) PCA is a very popular and versatile technique that is used to explain the variability within a data set (Amigo *et al*., 2015). The hypercube is separated into scores and loadings, which are then used to explain how similar characteristics are distributed within a sample. PCA provides an overview of the main differences in a sample and classification algorithms can now be used to predict constituents in an unknown sample (Geladi *et al.*, 1989). Many classification algorithms both supervised (e.g. Linear discriminant analysis (LDA), partial least squares regression (PLS-DA), k-nearest neighbours (kNN)) as well as unsupervised (e.g. k-means) exist.

NIR-HSI has shown its potential for detecting and identifying bacteria on a variety of sample matrices. It has been used to study lactic acid bacteria (LAB) on sliced, cooked ham (He et al, 2014; Foca *et al*., 2016), to determine *Escherichia coli* contamination on chicken fillets (Feng *et al*., 2013; Cheng & Sun, 2015), to detect damage caused by the *Pseudomonas* species on mushrooms (Gaston *et al*., 2011) as well as in complex liquid media e.g. milk and juice (Oliver *et al*., 2005). Food samples are however heterogeneous and very complex in nature, therefore, to solve this issue, a variety of studies were conducted using solid growth media to identify microorganisms. Yao *et al*., (2008), used Potato dextrose agar (PDA) to detect toxigenic fungi present in a variety of maize products. Yoon *et al*., (2010), investigated Campylobacter species grown on blood agar and Campy-cefex agar, to eliminate the confluent growth of *Campylobacter* species from non-*Campylobacter* species. To determine the effect of regions of interest (ROI) and spectral pre-processing on

the bacterial spectra, Windham *et al*., (2012) used sorbitol-McConkey (SMAC) agar to detect non-0157 Shiga-toxin (STEC) producing *E. coli*. In a similar study by Yoon *et al*., (2013), non-0157 Shiga-toxin producing *E. coli* was also investigated and was grown on Rainbow agar to differentiate between different serogroups.

Bacterial cells are sensitive to small changes such as cell cycle, culture conditions and sample preparation (Windham, *et al*., 2012). Various plating techniques have been implemented to detect and differentiate foodborne pathogenic bacteria. Yoon *et al*., (2010), investigated a spot plate technique to differentiate between *Campylobacter* and non-*Campylocater* species and achieved 97-99% classification accuracy in the 400-900nm wavelength range. In a follow up study conducted by Yoon *et al*., (2013), the researchers implemented a spread plate technique and used the spot plate data to validate the spread plates. However, found that the calibration model for the spot plates was sensitive to variations observed in the STEC populations as well as their growth conditions. The researchers ascribed this to the lack of sufficient spectral and spatial sampling of various colony populations the calibration model was provided with.

Kammies *et al*., (2016), evaluated the potential of the technique in the NIR range (780 – 2500 nm) to investigate whether it could be used to detect differences in foodborne pathogenic bacteria, based on their Gram-stain reactions and pathogenicity. The researchers were able to distinguish between Gram-positive and Gram-negative bacteria, however the study employed only one plating technique (streak plate), which could potentially influence the spectra obtained. Furthermore, while many forms of growth media exist, the researchers only used Luria-Bertani agar, not accounting for different variations of growth media. Bacterial growth is based on the environment they are placed in and could therefore potentially generate different spectral profiles.

Thus, the aim of this study was to investigate the potential of NIR-HSI as a rapid screening technique for the differentiation of foodborne pathogenic bacteria on different culture media. This was achieved by:

- determining the effect different culture media has on the NIR spectra of microorganisms
- investigating the effect of different plating techniques on the NIR spectra of microorganisms

1.1 REFERENCES

Aluwong, T., & Bello, M., (2015). Emerging diseases and implications for Millennium Development Goals in Africa by 2015 – an overview. *Veterinaria Italiana*, **46**(2): 137-135.

Amigo, J. M., (2010). Practical issues of hyperspectral imaging analysis of solid dosage forms. *Analytical Bioanalytical Chemistry*, **398**:93–109.

Amigo, J.M., Babamoradi, H. & Elcoroaristizabal, S. (2015). Hyperspectral image analysis. A tutorial. *Analytica Chimica Acta,* **896,** 34-51.

Baldauf, N.A., Rodriguez-Romo, L.A., Männig, A., Yousef, A.E., & Rodriguez-Soana, L.E., (2006). Effect of selective growth media on the differentiation of *Salmonella enterica* serovars by Fourier-Transform Mid-Infrared Spectroscopy. *Journal of Microbiological Methods*, **68**, 106–114

Bhardwaj, N., Bhardwaj, S.K., Nayak, M.K., Mehta, J., Kim, K-H., & Deep, A., (2017). Fluorescent nanobiosensors for the targeted detection of foodborne bacteria. *Trends in Analytical Chemistry*, Pp. 120-135.

Blackburn, C. de W., (2006). Food spoilage microorganisms. Woodhead Publishing Limited. Cambridge, England, Pp. 3-19.

Cen, H. & He, Y. (2007). Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends in Food Science & Technology,* **18,** 72-83.

Cheng, J-H., & Sun, D-W., (2015). Rapid and non-invasive detection of fish microbial spoilage by visible and near infrared hyperspectral imaging and multivariate analysis. *Science Direct*, **62**(2), 1060-1068.

Feng, Y.-Z., ElMasry, G., Sun, D.-W., Scannell, A.G.M., Walsh, D. & Morcy, N. (2013). Near-infrared hyperspectral imaging and partial least squares regression for rapid and reagentless determination of Enterobacteriaceae on chicken fillets. *Food Chemistry*, 138, 1829-1836.

Foca, G., Ferrari, C., Ulrici, A., Sciutto, G., Prati, S., Morandi, S., Brasca, M., Lavermicocca, P., Lanteri, S. & Oliveri, P. (2016). The potential of spectral and hyperspectral-imaging techniques for bacterial detection in food: A case study on lactic acid bacteria. *Talanta*, **153**, 111–119.

Gaston, E., Frias, J.M., Cullen, P.J., O'donnell, C. & Gowen, A. (2011). Hyperspectral imaging for the detection of microbial spoilage of mushrooms. Rapid and real-time prediction of lactic acid bacteria (LAB) in farmed salmon flesh using near-infrared (NIR) hyperspectral imaging combined with chemometric analysis. *Food Research International*, **62**:476-483.

Geladi, P., Isaksson, H., Lindqvist, L., Wold, S. and Esbensen, K. (1989). Principal component analysis of multivariate images. *Chemometrics and Intelligent Laboratory Systems, **5**,* 209-220**.**

He, H-J., Sun, D-W., Wu., D. (2014). Rapid and real-time prediction of lactic acid bacteria (LAB) in farmed salmon flesh using near-infrared (NIR) hyperspectral imaging combined with chemometric analysis. *Food Research International*, **62**, 476 - 483.

Kammies, T., Manley, M., Gouws, P. A. & Williams, P. J., (2016). Differentiation of foodborne bacteria using NIR hyperspectral imaging and multivariate data analysis. *Application of Microbiology and Biotechnology*, **100**: 9305–9320.

Law, J.W.-F., Ab Mutalib, N.-S., Chan, K.-G. & Lee, L.-H. (2015). Rapid methods for the detection of foodborne bacterial pathogens: Principles, applications, advantages and limitations. *Frontiers in Microbiology,* **5,** 770.

McClure, W.F. (2003) The Giant is running strong. *Journal of Near Infrared Spectrosc*opy, **11**: 487–518.

Oliver, S.P., Jayarao, B.M., & Almeida, R.A., (2005). Foodborne pathogens in milk and the dairy farm environment: Food safety and public health implications. *Foodborne pathogens and disease*, **2**(2): 115-129.

Wang, Y. & Salazar, J.K., (2016). Culture-independent rapid detection methods for bacterial pathogens and toxins in food matrices. *Comprehensive Reviews in Food Science and Food Safety*, **15**: 183-205.

Wang, K., Pu, H., and Da-Wen, S., (2018). Emerging Spectroscopic and Spectral Imaging Techniques for the Rapid Detection of Microorganisms: An Overview. *Comprehensive Reviews in Food Science and Food Safety*, **17**: 256-273.

Williams, P.J., & Kucheriavskiy, S., (2016). Classification of maize kernels using NIR hyperspectral imaging. *Food Chemistry*, **209**: 131-138.

Windham, W., Yoon, S-C., Ladely, S.R., Heitschmidt, J.W., Lawrence, K.C., Park, B., Narrang, N., Cray, W.C., (2012). The effect of regions of interest and spectral pre-processing on the detection of non-0157 Shiga-toxin producing Escherichia coli serogroups on agar media by hyperspectral imaging. *Journal of Near Infrared Spectroscopy*, **20**(5): p. 547-558.

World Health Organization, (2015). WHO estimates of the global burden of foodborne diseases: [www document] URL https://apps.who.int/iris/bitstream/handle/10665/199350/9789241565165_eng.pdf. September 2019.

Yao, H., Hrusk, Z., Kincaid, R., Brown, R.L., & Cleveland, T.E., (2008). Differentiation of toxigenic fungi using hyperspectral imagery. *Sensing & Instrumentation for Food Quality and Safety*, **2**:215–224.

Yoon SC, Lawrence KC, Line JE, Siragusa GR, Feldner PW, Park B, Windham WR (2010). Detection of *Campylobacter* colonies using hyperspectral imaging. *Sensing & Instrumentation for Food Quality & Safety*, 4(1):35-49.

Yoon, S.-C., Windham, W.R, Ladely, S.R., Heitschmidt, J.W., Lawrence, K.C., Park, B., Narang, N. and Cray, W.C., (2013). Hyperspectral imaging for differentiating colonies of non-O157 Shiga-toxin producing Escherichia coli (STEC) serogroups on spread plates of pure cultures. *Journal of Near Infrared Spectroscopy*, **21**(2): p. 81-95.

Zhao, X., Lin, C.-W., Wang, J. & Oh, D.H. (2014). Advances in rapid detection methods for foodborne pathogens. *Journal for Microbiology & Biotechnology,* **24,** 297-312.

CHAPTER 2

Challenges associated with NIR-HSI for the detection of foodborne pathogens on growth media: A review

2.1 INTRODUCTION

Food safety remains a global issue for public health (Jury and Vaux, 2005; Park *et al*., 2015; Ramírez-Castillo *et al.*, 2015). The World Health Organization (WHO) reported that 92 million deaths are related to foodborne illnesses annually, on the African continent alone (McDonnell, 2019). In the context of constant demographic growth, an additional threat is posed to agricultural systems as the need to feed the population is placed above the need for food safety. The drastic need for implementation of adaptive forms of comprehensive food security management is therefore continually on the rise. The increasing demand to ensure safe food supplies, calls for continuous development of rapid methods to detect foodborne pathogens (Mandal *et al.*, 2011). However, the ability to identify and establish metabolic functions of microorganisms in a sample remains a significant challenge for environmental and public health as physical differences play a major role in identification (Ramírez-Castillo *et al.*, 2015). The aim of developing these methods are ultimately to create fast, convenient and reliable systems. When used in combination with current conventional food microbiological methods, it will greatly enhance the sensitivity thereof (Mandal *et al.*, 2011). However, the problem is that although traditional molecular techniques have shown to improve the characterisation of these pathogens, there are several limitations linked to them (Ramírez-Castillo *et al.*, 2015). Challenges such as the lack of standardised protocols and laborious sample processing are ones that cannot be ignored any longer. Improving existing technologies to overcome these challenges will greatly increase safe food production which is a critical aspect for the food processing industry (ElMasry *et al*., 2012).

The implications of possible biological hazards at all stages of the food chain directly impacts food production systems and ultimately food security and food safety. Nearly a quarter of the damage caused to the global food supply is as a result of microbiological activity (Wang *et al.,* 2018). It is also becoming increasingly important to educate and train consumers on the issue of safe food as it is a global challenge. Studies have shown that consumers do not have sufficient knowledge in terms of handling food properly at home, storing food and preparing food safely (Unusan, 2005). The main contributing factors to reported outbreaks are contaminated raw foods, improper cooking and consuming foods from unsafe sources. Although different consumers exhibit different attitudes with regards to this concept, the reality remains that 'absolute safety' is just not possible (Wilcock *et al*., 2004). However, food safety incidences can be minimised and prevented with the help of rapid methods and the number of studies based on the development of these methods are on the rise (reviewed by Gowen *et al*., 2015; Wang *et al*., 2018). Wang & Salazar (2016), reported that 31 known pathogens are responsible for foodborne illnesses. It is also reported that among these

Campylobacter spp., *Listeria monocytogenes*, *Salmonella* spp., *Staphylococcus aureus*, *Clostridium perfringens*, *Vibrio* spp. and *Escherichia coli* (O157:H7) including strains of *E. coli* that produce Shiga toxins are the leading cause for foodborne related illnesses.

Food safety is defined as a concept contrasting that of food risk (Unusan, 2005). According to WHO, it is also referred to as the level of confidence that a specific foodstuff will not cause a consumer to be subjected to an illness because of preparing, serving or consuming that food substance as intended. Subsequently, a food risk is based on the assumption that a contaminated food product, that is likely to be consumed, poses an inherent hazard already present in the product (Tang & Wong, 2006). Foodborne diseases are normally a result of handling food in an improper manner as well as improper preparation of food in the consumers home (Scott, 2003). Consumers homes, albeit the last link to prevent foodborne illnesses, are a critical link in the chain before consumption. With that being said, contamination can occur at any stage of the food chain and preventable measures need to be implemented at all stages to ensure food quality and safety. Food related illnesses are completely preventable in most cases, should the correct safety procedures be implemented and followed from production to consumption (Unusan, 2005). The necessity to incorporate more effective microbiological monitoring, pathogen detection and health risk assessment is the force behind rapid method development (Mandal *et al.*, 2011).

A rapid method is defined as one that aims to address time delays associated with traditional culture-based techniques or an enhancement of an existing technique. The aim thereof is to aid in early detection and enumeration of microorganisms. Traditional techniques are usually based on cultivating bacteria grown from pure cultures and making use of viable counting measurements. Such techniques normally involve long incubation periods, various plating techniques and the use of different culture media (Mlynáriková *et al.*, 2015). A presumptive positive result is only obtained after enough time is afforded for the microorganisms to multiply to visible colonies taking up to three days (Park *et al*., 2015). A process known as fixation is then used prior to staining, to keep both internal and external cell structures of bacteria intact (Willey *et al.,* 2011). It prevents enzymes that has the potential to change the cell morphology, by inactivating it and therefore preventing it from being altered during observation. Two types of fixation procedures are generally used, namely heat fixation and chemical fixation. The first type is routinely used when working with bacteria and archaea and is done by smearing a thin film of cells onto a microscopic glass slide and then passing it through a flame. This procedure ensures that the overall morphology is preserved. The second type, however, preserves both the overall cell morphology as well as protect subcellular features. This is as a result of chemical fixative agents that react with finer cellular components e.g. proteins and lipids in order to inactivate them by penetrating the cells. This procedure is used for microorganisms of a larger volume and that are more fragile typically those that are at least 0.5 cm thick. After fixation, differential staining techniques are employed to identify bacteria based on their cell morphology.

It is evident that such practices are time consuming and laborious and it is therefore unavoidable to deny the time delays that it is usually associated with (Mandal *et al.*, 2011). These delays could potentially

lead to severe illnesses and fatalities, thus extensive research is constantly being done on early detection and identification techniques targeting foodborne pathogenic bacteria, to ultimately combat them. These techniques should preferably require minimal sample preparation, allow for a series of samples to be automatically analysed, permit rapid characterization against a known database and ensure ease of use at low costs.

Technologies that are currently in use for bacterial identification need to be improved in such a way that causative agents are detected accurately and more rapidly. Known spectral data can then be used to characterize microbial communities, which in turn reliably predicts pathogens contamination (Ramírez-Castillo *et al.*, 2015). In doing so, the creation of spectral libraries with accessible data is possible and therefore enhances the knowledge of foodborne pathogens. This increases the possibilities to predict contamination caused by pathogens and ultimately aids in the protection of public health. Furthermore, in order to optimise food safety and security standards and guidelines that are internationally recognised (e.g. Codex Alimentarius), such research is imperative for governmental control in order to be in line with national research priorities (Wang *et al.*, 2018b).

This review aims to address some of the potential challenges encountered with the application of NIR-hyperspectral imaging for microbiological analysis. The NIR spectra of bacteria could potentially be influenced by the culture media used as well as the way the sample is presented (plated out) to the growth medium. Following the optimization of the technique, it can be used as an alternative spectroscopic tool in the laboratory to enhance the advantages provided by traditional culture-based methods as well as address some of the short comings.

2.2 VIBRATIONAL SPECTROSCOPY

Various spectroscopic techniques have recently been applied to rapidly detect foodborne pathogens (Wang *et al.,* 2018). Vibrational spectroscopic techniques are being investigated for its potential to detect and identify microorganisms without destroying the sample. Furthermore, these techniques address the drawbacks experienced when applying traditional techniques, such as sample preparation and online monitoring of big sample volumes. The application of both Infrared (IR) and Fourier transform infrared (FT-IR) spectroscopy as a detection and classification method for microbiological cells has been carried out in previous studies (Windham *et al.*, 2012). Complete and highly characteristic spectral “fingerprints” of biological structure are obtained when applied to intact microbial cells. Combining these fingerprints with chemometrics, we can detect and classify different types of organisms.

2.2.1 Near infrared (NIR) Spectroscopy

NIR is a powerful analytical tool and has been widely used in the agricultural, petrochemical, pharmaceutical, clinical and environmental sectors and operates in the near infrared region between 780 – 2500 nm (Blanco & Villarroya, 2002). Overtones and combination modes of molecular vibrations modulating the dipole moment in organic molecules results in absorption spectra in this region (Šašić & Ozaki, 2010). The fundamental bond vibrations, such as stretching and bending, causes overtones to combine and results in an energy absorption (Manley, 2014). The resulting spectra contains important information about different functional groups, which all absorb at different wavelengths. Hydrogen bonding greatly affects the spectra in the NIR region, therefore molecules containing hydrogen bonds will provide an NIR spectrum (Šašić & Ozaki, 2010). The highest fundamental energies are contained by X-H bonds, where 'X' indicates C, O and N functional groups, thus the NIR spectra of organic molecules is dominated by their overtones and combination bands (Manley, 2014). NIR spectra are very complex due to the large numbers of possible vibrations. This results in overlapping of certain wavebands referred to as multicollinearity, resulting in peak broadening (Manley, 2014). Visually interpreting the NIR spectra, analysing chemical components and assigning features to them or extracting useful information from the spectra, has therefore proven to be a difficult task. Depending on the purpose of the analysis, appropriate regression techniques, linking absorption values to particular wavelengths and reference values of the constituents in question, therefore must be applied (Manley, 2014).

There are a few challenges encountered when applying this technique for microbiological systems. An NIR spectrum is a result of multiple data points that have been averaged, causing the loss of spatial localization (Vidal & Amigo, 2012). For this reason, not all constituents in a target sample are equally presented (McClure, 2003). Additionally, although it has been used to detect concentrations below 0.1%, it is insensitive to some minor constituents. This sensitivity has been improved by combining NIR spectroscopic techniques with wet chemistry, however the impact harsh chemicals has on the environment, may be a reason for a decline in the use thereof. Another major disadvantage is that NIR spectroscopy heavily relies on empirical quantitative models and is dependent on complex calibration procedures that involves complicated chemometric data analysis techniques (McClure, 2003). Each constituent requires to be calibrated and therefore requires the necessary software to perform these techniques as well as the appropriate hardware to be transferred from one instrument to another. In order to obtain satisfactory and reproducible results, careful monitoring is important. Despite these challenges however, NIR spectroscopy remains a viable analytical tool especially for on-line applications as the advantages far outweighs the shortcomings (McClure, 2003).

Although a scarce technique in the microbiological field, NIR spectroscopy has been used by Macaloney *et al.,* (1994) to report on *E. coli* fermentation by monitoring biomass and glycerol (Windham *et al*., 2012). The potential of the technique to quantify bacterial contamination on shredded cabbage was also

investigated by Suthiluk *et al.,* (2007). Another study was done to observe the changes in absorption and scattering properties during the growth phase of *Bacillus subtilus* in liquid culture (Dzongova *et al.*, 2009). Alexandrakis *et al.,* (2008) reported on the detection of bacteria in an isolated system. The same investigators also used NIR spectroscopy and Fourier transform mid-infrared on intact chicken breast muscle to detect spoilage (Alexandrakis *et al.,* 2012).

NIR spectroscopy is used to acquire molecular information by providing one spectrum of the target microorganism, however, does not provide any spatial resolution (Windham *et al.,* 2012). It is possible to perform analysis directly from agar plates or transferring it into small suspensions of approximately 20-200 µL or preparing smears onto microscopic glass slides. To obtain sufficient biomass for analysis, multiple coatings are often required. The technique provides both optimal sensitivity and selectivity, but lacks spatial information such as the amount present, shape as well as size of the colonies, which is very important when working with culture-based direct plating methods.

2.2.2 NIR Hyperspectral Imaging (NIR-HSI)

One way to address some of the disadvantages associated with NIR spectroscopy e.g. the lack of spatial dimension, is to combine it with imaging technology, specifically Hyperspectral imaging (HSI). HSI can be carried out in a variety of modes, which include reflectance, transmission as well as transflectance mode (Gowen *et al.,* 2015). Goetz *et al*., (1985) first investigated the application of hyperspectral imaging for remote sensing but it has since been applied in a variety of fields (e.g. agricultural, medical, pharmaceutical, environmental, geological) (ElMasry *et* al., 2012; Wang *et al*., 2018). The term 'hyperspectral', refers to a spectral data set that is multidimensional in character, which typically contains over hundreds of wavelength channels (Goetz *et al.,* 1985; Burger, 2006). Multispectral images refer to data that was previously acquired in just four to seven spectral bands and that contains channels from additional colour bands (Burger, 2006). Today, it is possible to acquire data in hundreds of spectral bands as made possible by imaging spectroscopy (Goetz *et al.,* 1985). The technique can therefore be used to overcome many other limitations linked to NIR spectroscopy because of its ability to provide very specific information about the target sample. The value thereof is prescribed to the ability it has to obtain a reflectance spectrum for each pixel in the image, making it especially suitable for heterogeneous samples. This provides us with an immense amount of position-referenced spectra meaning the question of what, where and how much can be answered.

A hyperspectral image is constructed by collecting spatial images through hyperspectral sensors and superimposing those images to create a three-dimensional data cube, known as a 'hypercube' (Huang *et al.*, 2014). A hypercube consists of two spatial (x, y) and one wavelength (λ) dimensions (Wang *et al.,* 2018). The chemical components of a sample can be determined using the unique spectral fingerprint of each pixel, displayed at each position within the sample. HSI can therefore be applied to successfully identify and

quantify the chemical components within a sample while simultaneously determining the distribution of each component within the sample.

2.3 INSTRUMENTATION

There are a few typical core components of any hyperspectral imaging system which include: light source, wavelength modulation system, detector and camera (Gowen *et al.,* 2015). Feng & Sun *et al*., (2012) refers to the imaging unit as the eye of the system, while the computer hardware and software functions as the brain. Different modalities require specific instrumentation depending on the purpose of the analysis. However, HSI technology is applied to most traditionally single point spectroscopic methods such as fluorescence, visible (Vis), infrared (IR), FTIR, NIR and Raman spectroscopy (RS) (Gowen *et al.,* 2015). For the purpose of this review, only NIR spectroscopy will be considered.

2.3.1 *Light sources*

Similar, to the role the human eye plays, the light source provides the system with light that interacts with target samples (Feng & Sun *et al*., 2012). Both chemical and physical information is contained in the detected portion, which is dispersed and projected onto a 2D detector array in an imaging spectrograph. A tungsten-halogen (TH) light source could be used as it is inexpensive and covers a broad range from 400 to 2500 nm (Gowen *et al.,* 2015). The drawback however is that it generates a significant amount of heat and may therefore cause changes or damage in the sample. This needs to be considered when imaging live microbial or food samples. Light emitting diodes (LED) are another source of light that is used, covering a wide spectral range, from visible to near-infrared region. Due to their small size, long lifespan and prevention of sample overheating, LED lights provide some advantage compared to TH bulbs. These lights currently covering the NIR range, however, are more expensive than TH sources.

2.3.2 *Wavelength modulation*

The light dispersion set-up is considered the core component of the HSI system. Imaging spectrographs, bandpass filter wheels and tuneable filters are available as wavelength dispersion devices (Gowen *et al.,* 2015). They are used to efficiently separate broadband incident lights into a variety of wavelengths.

2.3.3 *Detectors*

The detector is optimised based on the wavelength range and the purpose of the analysis. There are three basic choices of cameras for the hyperspectral imaging system detector (Huang *et al.,* 2014). HSI systems typically uses cameras that have a charge coupled device (CCD) or complementary metal oxide

semiconductor (CMOS) sensors in the visible near infrared region (Vis-NIR (400 – 1000 nm)) (Gowen *et al.,* 2015). These are both sensitive in the 300 – 1000 nm wavelength range but experiences an intense reduction in quantum efficiency below 400 nm and above 900 nm. The image acquisition time can be decreased by employing electron multiplying CCD (EMCCD) detectors as they provide a refined signal to noise ratio (SNR). This is important for heat sensitive microbial samples that are prone to photodegradation when exposed to the light source for long periods of time.

2.3.4 Image acquisition for HSI

Obtaining HSI data can be achieved using the three most common camera configurations as described below:

a) *"Whiskbroom" or "point mapping"*: The sample is scanned, point by point along two spatial dimensions (*x* and y). This is done by either moving the sample or detector to acquire a single spectrum for each pixel in the image (Qin *et al*., 2013).

b) *"Push-broom" or "line scanning"*: The sample is moved on a translation stage and the detector collects images by sweeping a line across the target sample, which generates spectral data (Šašić & Ozaki, 2010). The technique is used in Vis-NIR HSI and is usually 100 times faster than the whiskbroom configuration (Gowen *et al.,* 2015). This configuration is usually optimal when implemented for conveyor belt systems.

c) *"Stare-down" or "staring face"*: The entire sample is scanned one wavelength at a time, within the camera's field of view (FOV) (Qin *et al*., 2013). The sample and detector remain stationary when using this configuration. This technique is regarded as the only "true imaging technique" of the three methods.

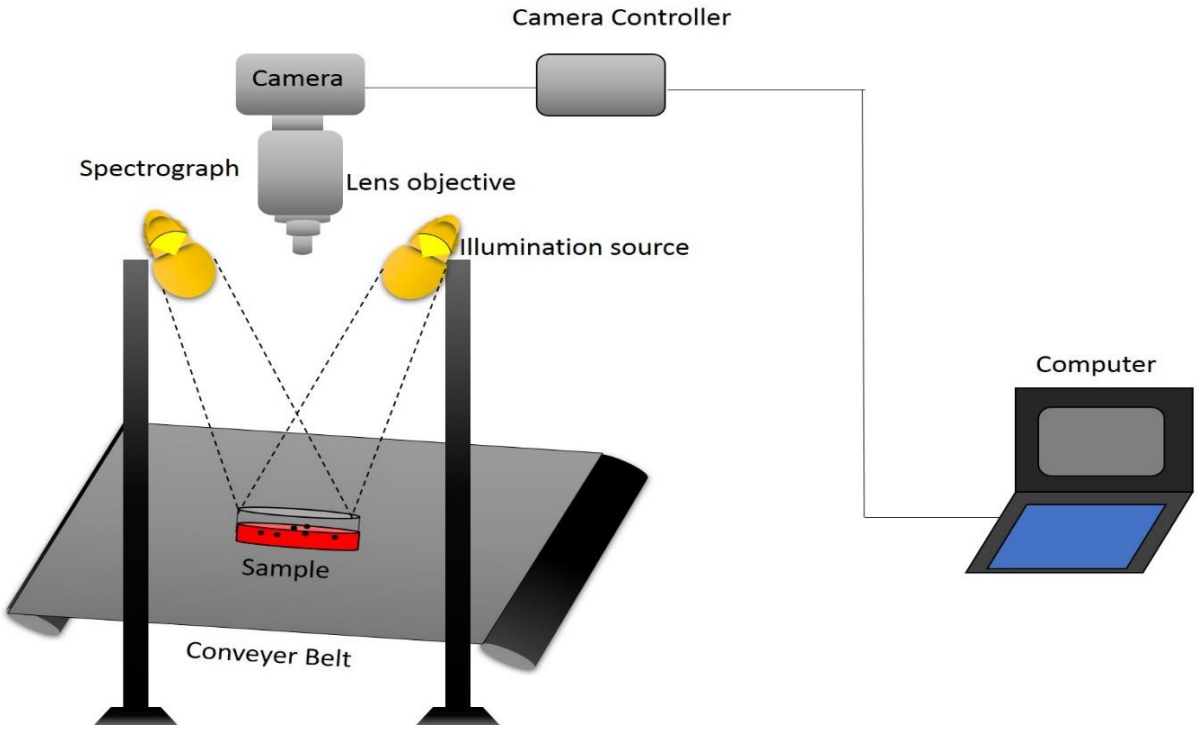

Figure 2.1 Typical configuration of an HSI system, illustrating the core components.

2.3.5 Multivariate Data Analysis (MDA)

The aim of applying multivariate data analysis (MDA) is to reduce the magnitude of data acquired with hyperspectral imaging, while extracting useful information in order to classify important areas of the target sample (Gowen *et al*., 2007). The most common mode of HSI is reflectance mode and is employed in the Vis-NIR (400-1000 nm), MIR (2500 – 25000 nm) or NIR (780 – 2500 nm) range. A variety of research areas such as food, agriculture, pharmaceutics and, most recently, microbiology has employed this technique.

The main steps involved include reflectance calibration, pre-processing, processing the data cube and lastly, image processing (Amigo, 2010). Spectral data acquired with spectroscopic techniques are often affected by noise generated by the instrument. Both signal to noise ratio (SNR) and baseline shifts are factors that need to be considered when working with NIR, MIR (due to light scattering) and RS (due to fluorescent influences) techniques. After sufficient data collection, it is crucial to apply pre-processing methods before any further processing is done to remove non-chemical biases and unwanted variation in the spectral data (Rinnan *et al*., 2009b). Cosmic radiation is a common occurrence in detectors used in HSI systems and causes dead or non-responsive pixels and spikes. This is often observed when using indium gallium arsenide (InGaAs) detectors that are typically used in NIR HSI systems. Identifying and removing these is required prior to any further processing. Smoothing methods such as polynomial baseline correction, median filtering, Savitzky-Golay (SG) derivative conversion (Savitzky & Golay, 1964) and unit variance normalisation methods, can be used to reduce noise in the spectra (Gowen *et al.,* 2015). Pre-processing techniques can also be applied in the spatial domain, meaning it can be processed both independently and in parallel (Fisher *et al*., 2003). Baseline correction methods such as standard normal variate (SNV), multiplicative scatter correction (MSC) and extended multiplicative scatter correction (EMSC) methods are often applied for NIR and MIR spectra (Amigo, 2010).

Hyperspectral images can be unfolded to produce a large spectral matrix (Burger *et* al., 2006). This matrix is composed of thousands or even millions of data points, which is amenable to standard chemometrics used for spectral data analysis (Amigo *et al*., 2013; Burger *et al*., 2006). MDA techniques (collectively known as chemometrics) are often applied in order to reduce the dimensionality of these data sets as they could contain bad pixels or redundant information (Amigo *et al*., 2013). The aim of applying these techniques are to extract useful information from the raw images without the loss of important information required to perform the analysis. Depending on the aim of the analysis, the data obtained is then used to either classify or quantify a specific target area in the sample. MDA techniques are categorised under either classification or regression and various algorithms for HSI evaluation have been proposed in the literature. The function and performance of the MDA depends on the aim of the analysis, the data set it is applied to and the quality of the operating system (Burger & Gowen, 2011). Spectral as well as spatial information contained in the hyperspectral images are used to classify objects by identifying comparable characteristics, whereas in regression modelling, the concentration of constituents in a target sample are predicted at pixel

level, making it possible to observe the spatial distribution/mapping of that particular component. Classification models are further grouped into supervised (e.g. k-nearest neighbours (kNN) or hierarchical clustering analysis (HCA)) and unsupervised (e.g. principal component analysis (PCA) or K-means) techniques, among others.

Analytical approach for microbial samples

Beebe *et al*., (1998), proposed six habits when applying chemometric methods. The first being able to examine the data after acquiring an appropriate data set. This is done to visually identify "obvious" errors by examining plots or tables containing the data. This will also indicate the need for pre-processing techniques and identify features that may need further investigation.

The second habit is to apply the identified pre-processing technique as required, as it is possible that random or systematic errors can mask the variation in question. Selecting appropriate pre-processing techniques generally requires understanding of the chemistry or physics that may be contained in these unwanted sources of variation. It is important to note that pre-processing techniques should be applied correctly, as it changes the data set and may remove some important information contained in the data. Foca *et al*., (2016), investigated lactic acid bacteria (*Lactobacillus curvatus* and *Lactobacillus sakei*) on sliced cooked ham in the 900 – 1700 nm spectral range. To remove the background from the Petri-dish, a combination of pre-processing techniques (SNV, detrending, first and second derivative, column mean-centering as well as autoscaling) were applied to evaluate the effectiveness thereof. Among these, detrending and column mean-centering provided the best contrast between the Petri-dish and background pixels.

The third habit involves examining the chemometric model by generating the model and diagnostics associated with it. The output results are dependent on the specific software being used and therefore different software packages will provide different results. Multivariate classification techniques (e.g. principal component analysis (PCA)) are used to statistically decompose datasets that are of a complex nature into structures that are easier to interpret (ElMasry *et al*., 2012). PCA, is an unsupervised analysis technique, considered to be a multifaceted tool used to recognize variation and trends within a dataset that may not be obvious through visual observation (Dorrepraal *et al*., 2016). It is performed on the two-dimensional data matrix, which is then reconstructed into principal components (PC's) and loadings. The loadings describe the relative importance or grouping of specific wavelengths in relation to each other.

After careful examination of the model, the model needs to be validated. Diagnostic tools are applied to each of the methods in order to assess the confidence that can be placed within the results obtained. Building models are simple but building reliable models are more complex. Chemical and mathematical knowledge is required to prove that the model is valid. If found that the model is not acceptable, the parameters used need to be re-evaluated. The second last step is to use this model for prediction by applying

the model to an unknown sample. This will produce the predicted properties or classification of the unknown sample(s). An example of this is Partial Least Squares (PLS) calibration models that are constructed to predict properties of future samples. Burger & Geladi (2006), used PLS regression to compare spectral results obtained from hyperspectral images to spectral results obtained by a FT-NIR instrument. The hyperspectral images showed promising results compared to that of the FTIR spectrometer. The researchers further highlighted the added spatial positioning provided by HSI, as substantially advantageous over conventional spectroscopic techniques.

The last and final step is to validate the prediction. One of the major advantages of applying multivariate techniques is to confirm the reliability of the model. A computer will always produce a prediction result if given a model and unknown variable. Validating the prediction model indicates the confidence that should be placed in the values predicted.

Reducing the dimensionality of data sets plays an integral role in image processing and a vast amount of data reduction methods are available. PCA is one of the most commonly used methods for reduction and exploratory purposes (Burger & Gowen, 2011). The general purpose of PCA is to transform the spectral variance contained in the hypercubes, to a smaller set of variables known as principal components. This is done by unfolding the 3-D hypercube, where each spectrum is represented by a row in the matrix. PCA is calculated as follows:

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E} \qquad \text{(eq. 2. 1)}$$

Where: **X** = Spectral Data

T= Score matrix

P' = Loading matrix

E = Residual matrix

After unfolding the hypercube and performing PCA (as would be done for regular spectral data), the scores are folded back into a 2-D score image (Qin *et al*., 2013).

The typical steps involved in analysing hyperspectral images are described below:

Reflectance calibration – the purpose of this step is to remove any background interferences that may occur during image acquisition. Also, to account for the dark current in the instrument as a response from the camera. A white reference standard is used to contrast the dark image (Nouri *et al*., 2013). The shutter is kept closed to acquire dark reference images (Nouri *et al*., 2013). The equation below depicts the calculation required for the corrected reflectance value (R):

$$\mathbf{R} = \frac{\mathbf{I} - \mathbf{D}}{\mathbf{W} - \mathbf{D}}$$

(eq. 2.2)

Where: **I** = Optical reflected signal (sample)

D = Dark response

W= Bright response (background)

Spectral Pre-processing - the aim of employing spectral pre-processing techniques is to minimize scattering effects due to inhomogeneities on the sample surface (Gowen *et al.,* 2015). These artefacts could arise as a result of instrument abnormalities that could generate spike points which is a sudden sharp point in the spectrum or unexpected missing or zero values (Vidal & Amigo, 2012). The shape and size of the sample also plays an important role as it will affect the amount and level of difficulty in removing the background. It is therefore important to select the appropriate region(s) of interest (ROI) by removing irrelevant background. This can be done by manual selection so only the desired area is included, by using histograms or by manually selecting the threshold value. Lastly, when a sample is irradiated, inherent artefacts such as light scattering or fluorescent interferences can occur. Pre-processing techniques include two groups: scatter correction methods and spectral derivatives (Rinnan *et al*., 2009a). When selecting the appropriate pre-processing technique, the goal is either to improve on a future analytical technique or bi-linear calibration model that will be applied (thus forcing it to obey Beer-Lambert's law) or to produce better results using a classification model. The Beer-Lambert equation (equation 2.3) suggests a linear relationship between the absorbance of the spectra and the concentration(s) of the constituent(s):

$$\mathbf{A = \varepsilon l c} \qquad \text{(eq. 2.3)}$$

Where: ***A*** *= Absorbance*

ε = Molar absorptivity

l = Path length

c = concentration

The most widely applied scatter corrective pre-processing techniques include: Standard Normal Variate (SNV) and Multiplicative Scatter Correction (MSC) (including Inverse MSC, Extended MSC (EMSC) and Extended Inverse MSC) (Rinnan *et al*., 2009a). Spectral derivation is *achieved by applying Savitzky-Golay (SG) and Norris-Williams (NW) smoothing techniques* (Savitzky & Golay, 1964; Norris 1959; Rinnan *et al*., 2009a).

Scatter correction methods:

We will consider the two most commonly used scatter correction techniques:

Multiplicative Scatter Correction (MSC)

MSC is a multi-wavelength concept used for optical correction. This is applied to separate light that is chemically absorbed from light scattering that occurs as a result of light interacting with particles in the sample (Geladi *et al*., 1985). The general purpose of employing this technique, is to remove or reduce

imperfections or non-linearities observed in the data (Rinnan *et al*., 2009b). There are two steps involved when applying MSC techniques:

The first is to account for additive and multiplicative contributions – estimation of the correction coefficients:

$$\boldsymbol{X_{org}} = \boldsymbol{b_0} + \boldsymbol{b_{ref,1}} . \boldsymbol{X_{ref}} + \boldsymbol{e} \qquad \text{(eq. 2.4)}$$

The second is to correct the recorded spectrum,

$$\boldsymbol{X_{corr}} = \frac{\boldsymbol{X_{org}} - \boldsymbol{b_0}}{\boldsymbol{b_{ref,1}}} = \boldsymbol{X_{ref}} + \frac{\boldsymbol{e}}{\boldsymbol{b_{ref,1}}} \qquad \text{(eq. 2.5)}$$

Where: $\boldsymbol{X_{org}}$ = one original sample spectra

$\boldsymbol{X_{ref}}$ = reference spectrum

$\boldsymbol{e}$ = un-modelled part of X_{org}

$\boldsymbol{X_{corr}}$ = corrected spectra

$\boldsymbol{b_0}$; $\boldsymbol{b_{ref,1}}$ = scalar parameters (differs for each sample)

Since spectra in the NIR region arise from differences in chemical composition, overlapping is a common occurrence due to highly correlating data (Rinnan *et al*., 2009a). It is therefore a difficult task to select one generic spectrum to represent the entire sample. Martens *et al*., (1983) therefore suggested that the calibration set average is used as a reference spectrum.

Standard Normal Variance (SNV)

The basic format for SNV is similar to that of the correction step for MSC (Rinnan *et al*., 2009a). In contrast to MSC, each wavelength is determined only on the wavelength axis and not making use of a reference spectra.

$$\mathbf{X_{corr}} = \frac{\mathbf{X_{org}} - \mathbf{a_0}}{\mathbf{a_1}} \qquad \text{(eq. 2.6)}$$

Where: $\mathbf{X_{corr}}$ *= corrected spectra*

$\mathbf{X_{org}}$ *= original sample spectrum*

$\mathbf{a_0}$ *= average value of sample spectrum*

$\mathbf{a_1}$ *= standard deviation of sample spectrum*

The drawback of SNV compared to MSC is that it does not involve a least-squares fitting step in the estimation of parameters and is therefore subjected to noisy spectra (Rinnan *et al*., 2009a). Despite MSC seemingly being a more robust approach, employing this technique should be done cautiously as it may result in additional outliers (Rinnan *et al*., 2009b; Gowen *et al*., 2015). Both techniques produce similar results and the following approximation is a representation of the similarities between MSC and SNV techniques (Dhanoa *et al*., 1994).

$$\mathbf{X_{MSC}} \approx \mathbf{X_{SNV}} \cdot \overline{\mathbf{s_X}} + \overline{\overline{\mathbf{X}}} \qquad \text{(eq. 2.7)}$$

Where: $\overline{\mathbf{s_X}}$ *= average standard deviation of all spectra*

$\overline{\overline{\mathbf{X}}}$ *= grand mean over all spectra*

} From raw/uncorrected

Spectral derivation methods

The aim of applying these techniques are to reduce/remove the multiplicative or constant additive effects as spectral baseline offsets and slopes (Rinnan et al., 2009a; Vidal & Amigo, 2012). Spectral derivatives include Savitzky-Golay (SG) polynomial derivative filters (Savitzky & Golay, 1964) and Norris-Williams derivatives (Norris, 1959). Both methods use a smoothing step before the derivative is determined to prevent a too large reduction in the SNR. First and second derivatives, in particular, are applied as a means of carefully selecting optimal wavelengths as well as removal of baseline shifts (ElMasry et al., 2012). SG and NW derivations do not typically produce similar results (Rinnan et al., 2009a).

Savitzky-Golay (SG) polynomial derivative

The smoothing step involved, is applied to the raw data by fitting a polynomial into a symmetric window generating the derivative at the centre point (Rinnan et al., 2009a). By applying analytical techniques to calculate the parameters for this polynomial, finding the derivative of any order of this function is greatly simplified. Two decisions need to be made: the window size which refers to the number of points used to calculate the derivative as well as the degree of the fitted polynomial and derivative order. Generally, spectral derivation amplifies the spectral features of the data but may cause an increase in noise levels (Vidal & Amigo, 2012).

Norris-Williams (NW)

The aim of this technique is to prevent an increase in noise in finite differences in the spectra (Rinnan et al., 2009a). Two steps are involved:

1. *Spectral smoothing – averaging a specific number of points, calculated as follows:*

$$\mathbf{X_{smooth,i}} = \frac{\sum_{j}^{m} = -m^{X_{org,i+j}}}{2m+1} \qquad \text{(eq. 2.8)}$$

Where: $\mathbf{m}$ = number of points in smoothing window

$\mathbf{i}$ = current measurement point in absorbance

$\mathbf{x}$ = number of samples

j = 'j^{th}' measurement point in absorbance

X_{org} = original sample spectra measured by the instrument

2. *The second step involves two approaches:*

- First order derivation - calculated by subtraction of two values with a given gap size larger than zero between them.
- Second order derivation – calculated by taking twice the smoothed value at a given point and taking smoothed values on either side at a gap distance between them.

- Calculated as follows:

$$\mathbf{x'_i} = \mathbf{x_{smooth,i+gap}} - \mathbf{x_{smooth,i-gap}}$$

$$\mathbf{x''_i} = \mathbf{x_{smooth,i-gap}} - \mathbf{2} \cdot \mathbf{x_{smooth,i}} + \mathbf{x_{smooth,i+gap}} \quad \text{(eq. 2.9)}$$

Classification – in order to extract meaningful features from the image, a variety of image analysis techniques can be applied after pre-processing (Huang *et al.,* 2014). Depending on the resolution of the image, spectral resolution and pixel binning, a magnitude of data can be acquired from each image. For this reason, multivariate analysis tools such as PCA, minimum noise fraction, partial least squares (PLS), linear discriminant analysis (LDA), wavelet transform, multi-linear regression (MLR), as well as artificial neural networks (ANN) among others are usually required for classification (Gowen *et al.,* 2007; Huang *et al.,* 2014). PLS classification, which is a supervised classification technique is the most commonly used method (Chevallier *et al.*, 2006). The aim of applying supervised techniques is building a calibration model that can be used to classify a variety of samples in future studies (Beebe *et al.,* 1998). A data set with known classification values is used to "train" computer software to recognize similar patterns in an unknown sample and uses it to classify it based on chemical differences. When PLS regression is used for discriminant purposes, it is known as partial least square discriminant analysis (PLS-DA) (Šašić & Ozaki, 2010). Two steps are involved in this process which firstly involves applying a PLS regression model on variables representing the different groups (pixels). The second step consists of using indicator variables to separate the groups into clusters and classes (Chevallier *et al.*, 2006; Šašić & Ozaki, 2010).

2.4 APPLICATIONS IN FOOD MICROBIOLOGY

HSI is a powerful analytical tool that has been applied in a vast variety of research fields (Gowen *et al.,* 2015). It is also currently being investigated as a candidate technology to be applied in microbiological applications. Microorganisms produce secondary metabolites that provides a characteristic fingerprint that can be used as an indication of microbial contamination of food and is the motivation for applying HSI for detection of microorganisms (Wang *et al*., 2018).

The bacterial cell wall

Bacteria are classed as single-celled organisms that contains the structural molecule peptidoglycan within their cell walls (Willey *et al*., 2011). Various studies investigated the biochemical structure of microorganisms (Al-Qadiri *et al*., 2006). Typically, a normal bacterial growth curve experiences four stages which include the lag, exponential log phase, stationary phase and death phase. These phases represent the typical timeline, bacterial colonies will experience when cultured in a complex or defined liquid medium. The nutrients supplied by the culture medium experiences a decrease as it is being used up over time. Although cells increase in size, no growth is shown during the lag phase, during log phase a linear increase in bacterial biomass is observed over time. When a zero-net increase in the number of viable bacterial cells is observed, the stationary phase has been reached. This means that the growth rate of bacterial cells is now equal to the death rate.

Obtaining pure cultures are therefore problematic and rapid methods (immunological, nucleic acid-based, biosensors, vibrational spectroscopic) are constantly studied to address time delays associated with it (Kammies *et al*., 2016). Some of these include those done directly on food matrices including fruit (Mehl *et al*., 2004; Vargas *et al*., 2005), vegetables (Liu *et al*., 2006; Siripatrawan, 2011) as well as meat for quality and grading purposes (Park *et al*., 2005, Qiao *et al* 2007; Peng *et al*. 2011, Kamruzzaman *et al*, 2012). In addition, studies have also been conducted on artificial substrates such as aluminium food cards (Dubois *et al*., 2005) as well as different growth media (Windham *et* al., 2012; Yoon *et al*., 2013a, 2013b; Kammies *et al*., 2016). Bacteria are either Gram-positive or Gram-negative, with their cell walls being one of the most important structural features (Willey *et al*., 2011). Gram-positive bacteria (e.g. *Staphylococci*) have a compact single cell wall that contains a very thick (about 20 - 80 nm) peptidoglycan (murein) layer and are prone to irregular cluster formation. Gram-negative bacteria, on the other hand, (e.g. *E. coli*) have a very complex cell-wall structure with a thin peptidoglycan layer (about 2-7 nm thick) which is covered by a thick membrane on the outside that is about 7 - 8 nm in diameter.

Bacterial differentiation is usually based on these differences, among others. The main contributing factors responsible for the differences between specific bacteria are proteins, lipids, carbohydrates, sugars, teichoic & lipoteichoic acids (Salton & Kim, 2018). Teichoic acids are secondary cell wall polymers (glycerol or ribitol) that contain phosphate groups that binds them in the cell walls of Gram-positive bacteria (Willey *et al*., 2011). Teichoic acids covalently bonded to the surface of the peptidoglycan layer are known as lipoteichoic acids. These acids are strongly negatively charged and therefore assist in giving the Gram-positive cell wall its negative charge. There are no teichoic acids present in Gram-negative cell walls. The periplasmic space is another major difference between Gram-positive and Gram-negative bacteria and is found between the plasma membrane and the cell wall. These differences are important when applying spectroscopic techniques such as NIR-HSI to successfully identify and differentiate bacterial pathogens that may be present in food matrices.

Employing HSI for contaminated food samples

Microbiological contamination in food is mainly caused by bacterial pathogens and the leading cause of food poisoning (Feng & Sun, 2012). Developing rapid methods to detect and identify these pathogens is becoming increasingly important. A number of studies have therefore been conducted to rapidly qualify and quantify pathogenic bacteria in a variety of food matrices.

Mehl *et al.,* (2004) investigated the potential of an HSI technique to detect defects and contaminants on apple surfaces. In the US, apples are an important agricultural commodity but is prone to *E. coli* contamination. Monochromatic images and second difference analysis (a method that characterises spectral differences on a broader spectrum in the 400 – 900 nm range) methods were used to examine spectral differences within the 430-900 nm spectral range. The second difference analysis method is commonly used in NIR analysis and can be used to construct symmetric (comparing each spectrum at each pixel to discriminate between weak and strong absorption bands) or asymmetric (comparing important spectral bands by providing a greater degree of freedom compared to the first method) models. Promising results were obtained achieving a coefficient of determination (R^2) of 0.99. A specific absorption band at 685 nm where chlorophyll absorbs, as well as two pigment absorption bands (722 and 869 nm) is used in the NIR region. The asymmetric second difference method is seen as a simpler, more general approach that does not require as much computation (e.g. PCA) as acquiring spectral band images directly and can therefore easily be implemented as a multispectral imaging technique.

Vargas *et al.*, (2005) conducted a similar study on cantaloupes using hyperspectral fluorescence imaging. Cantaloupes were artificially contaminated with a diluted bovine faecal solution and hyperspectral images were collected in the 425-774 nm range and contrasted to excitation in the ultraviolet (UV) range (320 – 400 nm). The entire hyperspectral data was used to perform PCA and clear distinction was observed between faeces spots and false positives in the second and fifth PC score images. The researchers found that the emission band at 675 nm exhibited the most favourable contrast between contaminated target and untreated areas. In the case where two bands were compared, a significant increase in contrast between the treated and untreated areas were evident. Relatively high detection rates were observed at 595/655 nm, 655/520 nm, and 555/655 nm which ranged from 79% - 96% across all faecal dilutions. However, several false positives were observed for the single band ratio's, which could be ascribed to physical damage on the sample surface.

Park *et al.,* (2005) evaluated the performance of an HSI system for identification of faecal and ingesta contaminated broiler carcasses in the 439-900 nm region. Sixty-four birds were measured and a prediction model for detecting surface contaminants was designed. Four different pre-processing methods were applied (uncalibrated raw data (uncalibrated without spectral smoothing) uncalibrated smooth data (uncalibrated with additional 20 nm spectral smoothing), calibrated raw data (percent reflectance calibrated with no spectral smoothing) and calibrated smooth data (percent reflectance calibrated with additional 20 nm

spectral smoothing), however the calibrated smooth method achieved the best results achieving 96.2% classification accuracy. Although the model achieved high accuracies, the researchers concluded that the samples used in the study was limited and that future studies should consider different sample conditions and different diets to determine the robustness of the system.

Another study was carried out in the 400-1000 nm range to assess *E. coli* contaminated fresh packaged spinach samples and raw vegetables (Siripatrawan *et al.,* 2011). Different initial concentrations of non-pathogenic *E. coli* K12 were inoculated into fresh spinach leaves. Reflectance spectra were obtained from different points on the target sample. Savitzky-Golay pre-processing with a second order polynomial was employed to remove any bad pixels from the acquired images. PCA was calculated to reduce the dimensionality of the spectral data. An artificial neural network (ANN) was trained with Bayesian regularization (to avoid overfitting) and used to construct a prediction map of the number of *E. coli* of each pixel in a sample. Control samples were successfully distinguished from the *E. coli* samples with a 98% cumulative variance and was attributed to metabolic compounds that could have been generated during incubation.

2.4.1 *Evaluating HSI for microbial samples on agar plates*

ElMasry *et al.*, (2012) suggests spatially locating and identifying a spoiled region on a food product, to fully exploit hyperspectral imaging capabilities. However, as food is of a complex matrix, separating and concentrating microorganisms for analysis can be challenging (Mandal *et* al., 2011). A variety of studies have been carried out on agar plates to address this issue.

Shiga toxin-producing *E. coli* (STEC) is a major health concern globally (Windham *et al*., 2012). In a study by Windham *et al*., (2012) six non-O157 STEC serogroups were evaluated and analysed directly on agar plates. The selected serogroups included O26, O45, O103, O111, O121 and O145 cell suspensions of each serogroup were grown on blood agar followed by inoculation onto Rainbow agar using a spot plate technique. Rainbow agar was selected as it is semi-transparent, and each serogroup will exhibit a specific colour on the agar surface. The data was pre-treated with Savitzky-Golay smoothing filter (window size 31, order of moment: 4) to reduce noise in the spectrum. Mahalanobis distance supervised classifier was employed to build classification models. The models achieved successful separation for serogroups O121 and O111 especially in the 550 – 700 nm wavelength range, with a classification accuracy of 80%. Serogroup O145 showed high absorptivity in the visible region and similar results were obtained from serogroups O26, O103 and O145 and therefore separation from other groups were possible.

More studies have been conducted to overcome the challenge of spatially locating microbiological contamination in a sample. One worth mentioning is a study by Dubois *et al.,* (2005) that used food specific cards to present the sample to the instrument. Strains of *Listeria innocua, Listeria monocytogenes, Bacillus cereus* (ATCC 700282), *Bacillus subtilis, Escherichia coli, Salmonella typhimurium* and *Salmonella entiritidis*

were suspended in brain heart infusion (BHI) and deposited into wells in the card. The samples were dried before imaging it in the 1000-2350 nm wavelength range. Successful differentiation of Gram-positive and Gram-negative bacteria was achieved based on their second derivative spectra, using the peak at 1940 nm (O-H stretch and O-H deformation). This was ascribed to the difference in their outer membranes that protects against dehydration. The band at 1680 nm, observed in the spectra of *B*. *cereus,* relating *cis*-double bonds in unsaturated fatty acids, produced image contrast between all the other bacteria. It was however not possible to differentiate between all the species using just one wavelength, therefore the authors developed a two-step PLS-DA model employing the full wavelength range which achieved excellent separation between all species.

2.4.2 *Studies conducted to address issues encountered when using HSI for bacterial pathogen detection*

Isolating pure cultures is a lengthy procedure but is important in order to obtain sufficient spectral data (Kammies *et al*., 2016). Various plating methods should therefore be explored to determine which technique would provide optimal results. A new screening technique was developed by Yoon (2010) and co-workers, using spot plates to detect *Campylobacter* colonies directly on agar plates in the 400 – 900 nm wavelength range. *Campylobacter* colonies can take up to 48 h to grow on agar media and often much difficulty is experienced when differentiating between *Campylobacter* colonies from non-*Campylobacter* colonies as they frequently grow together. Campy-Cefex and blood agar were used as culture media, the results however showed that the blood agar produced the best results in terms of *Campylobacter* detection accuracy. It was found that the bacteria grown on blood agar achieved 97-99% detection accuracy provided by a two band (426 and 458 nm) algorithm. This highly accurate method has proven to have ground-breaking implications for the early detection of *Campylobacter* colonies as well as detection for other pathogenic bacteria.

Solid growth media could potentially generate differences in spectral responses (Kammies *et al*., 2016). It is therefore important to control this variable and to determine the suitability of specific media for specific microorganisms. Williams *et al.,* (2012) used NIR-HSI to study three strains of Fusarium spp. of each of the following: *Fusarium subglutinans, Fusarium proliferatum and Fusarium verticillioides*. The fungi were incubated for either 72 or 96 h and then streaked out onto potato dextrose agar (PDA) in Petri dishes. A combination of pre-processing techniques (including SNV, MSC and SG) were applied to the data set but provided no improvement in the spectral data. Using PLS-DA, it was found that *F. verticillioides* was the least correctly predicted (between 16 – 47%). Better prediction results were obtained for *F. subglutinans,* with 78 – 100% prediction accuracy and *F. proliferatum* achieved 60 – 80% correctly predicted pixels. The inhomogeneous growth of the Fusarium spp. colonies made it ideal for NIR-HSI analysis but could not be separated by visual observation based on the shape and colour. The PLS-DA model achieved successful discrimination between *Fusarium subglutinans* and *Fusarium proliferatum* but was not able to distinguish between *Fusarium verticillioides* and other species. This was attributed to the large amount of spectral

variation observed within this class. The researchers concluded that future studies should be focussed on employing NIR-HSI for fungal discrimination using different growth media and taking temperature and water activities into account.

It is important to note that some of the studies mentioned above were conducted in the Vis-NIR (400 – 1000 nm) region. This implies that classifications were made on the basis of colour variations between the bacterial colonies (Kammies *et al*., 2016). To answer whether NIR-HSI could be used to accurately detect pathogenic bacteria beyond the visible region, Kammies *et al*., (2016) used a pixel-wise approach coupled with multivariate analysis. The spectral data was obtained in the 920-2514 nm wavelength range and PCA was calculated for individual images. The PCA score plots showed the chemical differences observed between colonies, which was visually difficult to distinguish. Successful separation for *B. cereus* from *E. coli* and *S. enteritidis* was possible along the first principal component (PC1), which explained 58.1 % of the variation, as well as for *E. coli* from *S. enteritidis*, along PC2 (7.75 %). The separation was ascribed to the variation in amino acid and carbohydrate content found in *S. enteritidis*, *B. cereus* and *S. aureus.* Distinction between both *S. aureus* and *S. epidermidis*, was also possible due to the formation of two distinct clusters in the PCA score plots. The researchers concluded that NIR-HSI could successfully distinguish between Gram-positive and Gram-negative, using bacteria from different species and within the same genera on solid growth media.

The above findings along with other recent studies including *Pseudomonas tolasii* contamination on mushrooms (Gaston *et al.*, 2011); determining total viable count and psychrotrophic plate count on pork meat (Barbin *et al*., 2013) and the differentiation of toxigenic fungi (Yao, *et al*., 2008) using NIR-HSI for the detection of foodborne pathogenic bacteria are summarised in Table 2.1.

Table 2.1 Recent advancements in hyperspectral imaging applications for rapid detection and differentiation of foodborne pathogens

Samples	Microorganisms	Growth media	Spectral range	MDA technique	References
Food specific cards	*Listeria innocua*, *Listeria monocytogenes*, *Bacillus cereus* (ATCC 700282), *Bacillus subtilis*, *Escherichia coli*, *Salmonella typhimurium* & *Salmonella entiritidis*	Brain Heart Infusion (BHI)	1000 – 2350 nm	PLS	Dubois *et al*., 2005
Campylobacter on animal products	*Campylobacter*	Blood agar & Campy-Cefex	400 – 900 nm	Spectral feature fitting and continuum-removed band ratio	Yoon *et al*., 2010
Fruit	*E. coli*	Apples (surface defects & contaminants)	430 – 900 nm	PCA	Mehl *et al*., 2004
		Cantaloupes (faecal contamination)	425 – 774 nm	PCA	Vargas *et al*., 2005
Vegetables	*E. coli*	Spinach (faecal contamination)	400 – 1000 nm	PCA, ANN	Siripatrawan *et al*., 2011
	Pseudomonas tolasii	*Agaricus bisporus* (mushrooms – distinguish between: wholesome, mechanically damaged & infected mushrooms)	445 – 945 nm	PLS-DA	Gaston *et al*., 2011

Meats		Chicken carcasses (faecal contamination)	439 – 900 nm	PCA	Park *et al*., 2005
	PPC & TVC	Porcine meat	900 – 1700 nm	PLS	Barbin *et al*., 2013
	E. coli	Chicken fillets	930-1450 nm	PLSR	Feng *et al*., 2013
Foodborne pathogens on growth media	Fungi	Potato dextrose agar (PDA)	400 – 1000 nm	ML	Yao *et al*., 2008
	Fungi	PDA	1000 - 2498	PCA, PLS-DA	Williams *et al*., 2012
	E. coli	Rainbow agar	400 – 1000 nm	Mahalanobis distance classifier	Windham *et al*.,2012
	E. coli (STEC)	Rainbow agar	368 – 1024 nm	PCA (Mahalanobis distance classifier & kNN)	Yoon *et al*., 2013
	Bacillus cereus, Staphylococcus aureus, Staphylococcus epidermidis, Escherichia coli, Salmonella enteritidis,	Luria-Bertani agar	920-2514 nm	PCA, PLS-DA	Kammies *et al*., 2016
	Fungi	PDA	1000 – 2498 nm	PCA, PLS-DA	Williams *et al*., 2019

Notes: Psychrotrophic plate count (PPC), Total viable count (TVC), Multiple linear regression (MLR), principal component analysis (PCA), artificial neural network (ANN), k-nearest neighbour (kNN), partial least squares (PLS), partial least squares discriminant analysis (PLS-DA), partial least squares regression (PLSR), soft independent modelling of class analogy (SIMCA), Maximum likelihood (ML).

2.5 FACTORS AFFECTING SPECTRA OF MICROBIOLOGICAL SAMPLES

Hyperspectral imaging techniques have a broad range of applications and its application to microbiological systems are on the rise. There are however a few challenges that is associated with the application of NIR-HSI to microbiological systems. A number of key challenges associated to microbiological applications using NIR-HSI can be summarised as follow.

2.5.1 *Plating techniques*

Microorganisms usually grow in complex mixed cultures when in their natural habitat (Willey *et al.,* 2011). This of course is very problematic as a single organism cannot be examined in a mixed culture. The use of various techniques is required to isolate pure cultures prior to analysis. One of the most commonly used methods for the preparation of pure cultures is the **streak plate** method. The procedure is essentially a dilution process that results in single colonies and eventually isolated pure cultures. Another common technique used to obtain pure cultures is the **spread** technique. A serial dilution is prepared before obtaining spatial separation. Using the spread plate method for food samples a small volume of ±25 to 250 cells is required to be transferred to the centre of an agar plate. The number of colonies formed are equal to the number of microbes in the sample that has the ability to grow in the specific medium used. It is therefore possible to use both described techniques for HSI application, but streak plates are preferred for the differentiation of foodborne pathogens as it provides uniformly formed single colonies (Kammies *et al.,* 2016). The chemical environment of the pathogens however is altered when using these plating techniques (Jiang *et al.*, 2004). This can therefore potentially affect the NIR spectra of the isolated bacterial cells. Investigating the effect, a specific isolation procedure has on the spectra is therefore of utmost importance.

2.5.2 Growth media

Few studies have been conducted in terms of specific growth media needed to obtain optimal growth environments for microorganism. A culture medium is described as a solid or liquid culture prepared to provide effective growth by providing the required nutrients, assist in transportation and storage of microorganisms (Willey *et al.,* 2011). Specialised media is an essential part in identifying and isolating foodborne pathogens. Depending on the organism being investigated, the medium is chosen based on its ability to provide sufficient sources of energy, carbon, nitrogen, sulphur, phosphorous as well as various other minerals. Both solid and liquid media are used in routine microbiological analysis, however solid media is of specific importance as it provides the ability for isolation of pure cultures (Willey *et al.,* 2011). **General purpose** or **supportive media** includes tryptic soy broth and tryptic soy agar and they are able to provide the growth requirements for a variety of microorganisms. **Enriched media** refers to supportive media that has

been fortified with blood (referred to as blood agar) or other nutrients to encourage growth of fastidious microbes. When aiming to grow specific organisms, the use of **selective media** is particularly useful as it is made to favour the growth of such organisms. In order to grow Gram-negative bacteria, bile salts or dyes is used as it inhibits the growth of Gram-positive bacteria and does not affect Gram-negative bacteria (Willey *et al.,* 2011). Examples of these are endo agar, eosin methylene blue agar and MacConkey agar among others. In order to differentiate between different groups of organisms, **differential media** is used. An example of a differential as well as an enriched media is blood agar, whereas mannitol salt agar (MSA) is an example of both a selective and differential agar. MSA is usually used for the growth of *Staphylococci* which uses mannitol as a source of carbon energy. It produces acidic products which causes phenol red that serves as a built in pH indicator to turn yellow (Willey *et al.,* 2011). Certain growth media has the potential to influence the microbial fingerprint and as the molecular composition of the cells are measured, the spectra may be influenced by the culture conditions (Mlynáriková *et al.*, 2015). The possibility thereof should therefore be investigated.

It is possible to analyse microorganisms directly from colonies that have been grown on agar plates, on solid growth media containing microcolonies, or from single cells suspended in a liquid media after being cultivated (Mlynáriková, *et. al.,* 2015). Bacteria are therefore only required to be cultured until formation of microcolonies which are defined as colonies that have grown up to six hours during incubation and that, depending on the organism, is 10-110 mm in diameter (Maquelin *et al.,* 2002). Mentioned previously Foca *et al.,* (2016), used gel cultures to detect bacteria on cooked ham samples. The researchers used a streak plate method to inoculate *Lactobacillus* onto De Man, Rogosa and Sharpe (MRS) agar and incubate the bacteria for 72 hours at 37 °C overnight in glass Petri-dishes. A clear separation of the bacterial samples from the gel using PCA was obtained. Previously mentioned Kammies *et al.,* (2016) used a streak plate method and LB agar for the differentiation of foodborne pathogens. A universal culture media was used to account for any spectral interferences that may occur as a result thereof.

Culture media differ in their composition based on the nutritive requirements of the organisms being analysed and can also serve as diagnostic or selective tool (Mlynáriková *et al.*, 2015). An integral part of bacterial differentiation and identification is to find suitable culture media for selected organisms. This is done by assessing nutrient requirements provided by specific culture media that is needed for specific organisms. In preparing and choosing adequate culture media based on the nutritional requirement of the organisms, the reproducibility of the method may be enhanced, and the data processing and evaluation may be simplified. Once the appropriate media is established depending on the bacteria used, microorganisms can further be analysed directly from colonies grown on these agar plates or by assessing single cells contained in a suspension of in liquid media.

2.5.3 Sample presentation

Culturing and enumeration methods are commonly recognized as the golden standard for the detection and identification of foodborne pathogens (Feng & Sun, 2012). These methods are however seen as destructive and laborious and NIR-HSI is therefore being proposed as a good candidate to greatly enhance these methods for rapid, accurate detection One of the main advantages of the technique, is that little to no sample presentation is required (Amigo, 2010). However, it is preferred that samples are flat with a uniform shape to avoid possible light scattering and that it is positioned parallel to the image plane taking the focussing nature of the device into account. Water interferences are another challenge observed in the NIR region and various sample presentations have been proposed to overcome issues that are brought about by signal interferences by water and low microbial concentrations (Gowen *et al*., 2015). Drying samples before imaging was found to be the best solution to overcoming water interferences. However, the drying process may bring about unwanted changes in the sample especially for biofilm formation. The use of novel microfluidic cells by some researchers, have proven to overcome this issue.

Another technique that enables imaging of intact microbial cells in their natural environment, is darkfield HSI. However, when imaging live cells due to cell motility, image blurring may occur. Park *et al*., (2011) developed a method using FT-IR to distinguish between live and dead cells of two *Salmonella* serotypes. Cell suspensions were loaded onto a zinc selenide (ZnSe) attenuated total reflection (ATR) crystal surface prior to scanning. PCA was calculated and the images were obtained using three wavelength ranges (900 – 1300 nm; 1300 – 1800 nm and 3000 – 2200 cm^{-1}). SIMCA was applied to validation sample sets and classification accuracies of 100% was obtained. Signature peaks of fatty acids, DNA, RNA, proteins and a combination of amide I and II regions were used to explain the main differences observed within both serotypes.

Dubois *et al*., (2005) developed food specific cards by first isolating target bacterial cells from various foods and then placing a thick film of bacteria into a well in the card. Furthermore, the researchers analysed images at specific wavelengths to carefully select a threshold that separates different bacterial strains from others and enables identification at species level instead of working with spectra. The researchers concluded that analysis should be done directly on food to reduce or omit the time taken to prepare the samples before imaging (Amigo, 2010). A study done by Siripatrawan *et al*., (2011) investigated *E. coli* on spinach leaves surfaces. Serial dilutions of stock suspensions of contaminated spinach leaves were prepared and imaged in the 400-100 nm range. To address time delays, the suspensions were inoculated onto specialized aerobic count plates containing Violet Red Bile agar (VRBA). This specialized media indicates glucuronidase activity in *E*. *coli,* enabling rapid identification prior to direct imaging of the samples.

2.5.4 Spatial resolution

Spatial resolution is identified as one of the key limiting factors in HSI applied to microbial samples (Gowen *et al*., 2015). The spatial resolution is used to establish the smallest cell or region to be examined. Vis-NIR systems provides a spatial resolution of approximately 300 µm which could result in some areas on the sample surface containing small colonies to go undetected. There are several factors that need to be considered including the physical material of the sample, its shape and size as well as the spectral radiation process (whether incident radiation is reflected or absorbed by the sample) (Amigo, 2010). The capabilities of the instrument in terms of pixel size is another important consideration as it may affect the information contained in the spectra obtained. It is therefore important to find a compromise between pixel resolution, the FOV imaging plane and the time taken to acquire images (Gowen *et al*., 2015).

2.5.5 Limit of detection (LOD)

The quantification of microbial cells at low concentrations is another challenge for NIR-HSI. A number of studies are done to investigate microbiologically contaminated food and to specifically identify microbial loads in food that are at volumes higher than 10^2 cfu/mL. Traditional plating methods may contain initial cell concentrations higher than the lowest level of bacteria that can be detected (Gowen *et al*., 2015). This is a limitation as applications may be limited to high microbial loads in terms of bacteria that has an impact on food quality and safety.(Buchannan & Deroever, 1993) The effectiveness of both microscopic and direct macroscopic analysis in NIR-HSI technology, should be investigated further to address this limitation.

Park and co-workers (2012), used Acousto-optic tunable filter (AOTF) hyperspectral microscope to characterise spectra of *Salmonella* and *E. coli* species. The goal of the study was to enhance the LOD to less than 100 cells. Spectra were therefore obtained from 100 cells of each species and compared. Six important wavelengths were obtained for the *Salmonella* species however, the researchers encountered some difficulty as only one intense peak was observed for *E. coli*. The dominant peaks observed in this study was ascribed to the structural differences of the two species and not as a result of intra-class differences.

2.6 CONCLUSION

The issue of food quality and safety is an ever-growing point of concern. Traditional methods are usually applied to identify and differentiate pathogens that are related to foodborne illnesses, however these techniques are known to be extremely time consuming with questionable sensitivity. Several challenges are linked to ensuring effective food safety and security systems as well as the implementation thereof. The need to improve food production systems to prevent or minimize food safety issues therefore exist and the focus of these systems is to minimize time delays and the reliability of such systems. To optimize food quality, food safety and security standards, such research is imperative.

NIR-HSI have been investigated for the detection of foodborne pathogens and successful methods along with multivariate techniques (PCA, PLS-DA, LDA & PLSR among the most common) have been proposed by many studies. However, limited work reported on the effect growth media has on the spectral fingerprint and not enough studies have been done modelling time series data from microbial communities. By developing a method that can successfully bring about structural identification of food-borne pathogens using NIR-HSI and assessing these challenges, several advantageous are observed. Furthermore, hyperspectral imaging provides an added spatial dimension and numerous studies have been done to reduce some of these challenges. The advantages of the technique such as speed, sensitivity, quantitative as well as qualitative analysis can potentially address any additional associated challenges. It has the potential to be used as a tool for microbiology to construct spectral libraries for future use.

2.7 REFERENCES

Alexandrakis D, Downey G, Scannell AG (2008) Detection and identification of bacteria in an isolated system with near-infrared spectroscopy and multivariate analysis. J Agric Food Chem 56(10):3431-3437.

Alexandrakis, D., Downey, G. & Scannell, A. G. M., (2012). Rapid Non-destructive Detection of Spoilage of Intact Chicken Breast Muscle Using Near-infrared and Fourier Transform Mid-infrared Spectroscopy and Multivariate Statistics. *Food Bioprocess Technology*, **5**: 338.

Al-Qadiri, H. M., Lin, M., Cavinato, A. G., Rasco, B. A., (2006). Fourier transform infrared spectroscopy, detection and identification of *Escherichia coli* O157:H7 and *Alicyclobacillus* strains in apple juice. *International Journal of Food Microbiology*, **111**: 73-80.

Amigo, J. M., (2010). Practical issues of hyperspectral imaging analysis of solid dosage forms. In: Anal Bioanal Chem, **398**:93–109.

Amigo JM, Martí I, Gowen A (2013) Hyperspectral imaging and chemometrics: A perfect combination for the analysis of food structure, composition and quality. Marini F (ed) Chemometrics in Food Chemistry. 1st edn. Elsevier, pp 343-370.

Barbin, D.F., Elmasry, G., Sun, D.-W., Allen, P. & Morsy, N. (2013). Non-destructive assessment of microbial contamination in porcine meat using NIR Hyperspectral Imaging. *Innovative Food Science & Emerging Technologies*, 17, 180-191.

Beebe, K.R., Pell, R.J & Seasholtz, M.B. (1998). Chemometrics: A Practical Guide. John Wiley & Sons, Inc., Canada, Pp. 1–8.

Blanco, M. & Villarroya, I. (2002). NIR spectroscopy: A rapid-response analytical tool. *Trends in analytical chemistry*, **21**(4): 240-250.

Buchanan, R. L. & Deroever, C. M. (1993). Limits in assessing microbiological food safety. *Journal of food protection*, **56**, 725 – 729.

Burger, J. (2006). Hyperspectral NIR image analysis. Swedish University of Agricultural Sciences Umeå.

Burger, J. & Geladi, P. (2006). Hyperspectral NIR image regression part ii: *Dataset pre-processing diagnostics. Journal of Chemometrics*, 20, 106-119.

Burger, J. & Gowen, A. (2011). Data handling in hyperspectral image analysis. *Chemometrics and Intelligent Laboratory Systems,* **108,** 13-22.

Chevallier S, Bertrand D, Kohler A, Courcoux P (2006) Application of PLS-DA in multivariate image analysis. *Journal of Chemometrics* 20(5):221-229.

Dhanoa, M., Lister, S., Sanderson, R. & Barnes, R. (1994). The link between multiplicative scatter correction (msc) and standard normal variate (SNV) transformations of NIR spectra. *Journal of Near Infrared Spectroscopy,* **2,** 43-47.

Dorrepaal, R., Malegori, C., Gowen, A. (2016). Tutorial: Time series hyperspectral image analysis. *Journal of Near Infrared Spectroscopy*, **24,** 89–107.

Dubois, J., Lewis, E.N., Fry F.S. & Calvey, E.M. (2005). Bacterial identification by near-infrared chemical imaging of food-specific cards. *Food Microbiology*, **22**(6): Pp. 577-583.

Dzhongova, E., Harwood, C. R., Thennadil, S. N., (2009). Changes in the Absorption and Scattering Properties in the Near-Infrared Region during the Growth of *Bacillus subtilis* in Liquid Culture. *SAGE Journals*, *Applied Spectroscopy,* **63**: 25 – 32.

ElMasry, G., Kamruzzaman, M., Sun, D & Allen, P., (2012). Principles and Applications of Hyperspectral Imaging in Quality Evaluation of Agro-Food Products: A Review. *Critical Reviews in Food Science and Nutrition*, **52**:999–1023.

Feng, Y.-Z. and D.-W. Sun. (2012). Application of hyperspectral imaging in food safety inspection and control: A review. *Critical Reviews in Food Science and Nutrition*, **52**(11): p. 1039-1058.

Feng, Y.-Z., Elmasry, G., Sun, D.-W., Scannell, A.G.M., Walsh, D. & Morcy, N. (2013). Near-infrared hyperspectral imaging and partial least squares regression for rapid and reagentless determination of Enterobacteriaceae on chicken fillets. *Food Chemistry*, 138, 1829-1836.

Fisher, R., Perkins, S., Walker, A. & Wolfart, E., (2003). Hypermedia image processing. [www document] URL http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/BMVA96Tut/node30.html. 06 September 2019.

Foca, G., Ferrari, C., Ulrici, A., Sciutto, G., Prati, S., Morandi, S., Brasca, M., Lavermicocca, P., Lanteri, S. & Oliveri, P. (2016). The potential of spectral and hyperspectral-imaging techniques for bacterial detection in food: A case study on lactic acid bacteria. *Talanta*, **153**, 111–119.

Gaston, E., Frias, J.M., Cullen, P.J., O'donnell, C. & Gowen, A. (2011). Hyperspectral imaging for the detection of microbial spoilage of mushrooms. Oral Presentation MCF1004 at the *11th International Conference of Engineering and Food. Athens*, Greece, May, 2011.

Geladi, P., Macdougall, D. & Martens, H. (1985). Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Applied Spectroscopy,* **39,** 491-500.

Goetz, A.F.H., Vane, G., Solomon, J.E. & Rock, B.N. (1985). Imaging Spectroscopy for Earth Remote Sensing. *Science,* **228**, 1147-1153.

Gowen, A.A., O'Donnel, C.P., Cullen, P.J., Downey, G., & Frias, J.M., (2007). Hyperspectral imaging: an emerging process analytical tool for food quality and safety control. *Trends in Food Science and Technology*, **18**(12): p. 590-598.

Gowen, A.A., Feng, Y., Gaston, E. & Valdramidis, V. (2015). Recent applications of hyperspectral imaging in microbiology. *Talanta*, **137**, 43-54.

Huang, H., Liu, L. & Ngadi, M.O. (2014). Recent developments in hyperspectral imaging for assessment of food quality and safety. *Sensors,* **14**, 7248-7276.

Jury, W.A. & Vaux, H. (2005). The role of science in solving the world's emerging water problems. *Proceedings of the National Academy of Sciences*, **102**, 15715–15720.

Kammies, T., Manley, M., Gouws, P. A. & Williams, P. J., (2016). Differentiation of foodborne bacteria using NIR hyperspectral imaging and multivariate data analysis. Application of Microbiology and Biotechnoly, **100**: 9305–9320.

Kamruzzaman, M., ElMasry, G., Suna, D. & Allen, P. (2012). Prediction of some quality attributes of lamb meat using near-infrared hyperspectral imaging and multivariate analysis. In: *Analytica Chemica Acta,* **714**: 57-67.

Liu, Y., Chen, Y.R., Wang, C.Y., Chan, D.E & Kim, M.S. (2006). Development of Hyperspectral Imaging technique for the detection of chilling injury in cucumbers; spectral and image analysis. *Applied Engineering in Agriculture*, **22**(1): 101-111.

Macaloney, G., Hall, J. W., Rollins, M. J., Draper, I., Thompson, B. G., McNeil, B., (1994). Monitoring biomass and glycerol in an *Escherichia coli* fermentation using near-infrared spectroscopy. *Biotechnology Techniques*, **4**: 281 – 286.

Mandal, P.K., Biswas, A.K., Choi, K. & Pal, U.K. (2011). Methods for Rapid Detection of Foodborne Pathogens: An Overview. *American Journal of Food Technology*, Pp.87-99.

Manley, M. (2014) Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials. *Chemical Society Review*, **43**(24): p. 8200-8214.

Maquelin, K. & Choo-Smith, L.P. (2002). Rapid identification of Candida species by confocal Raman microspectroscopy. *Journal of clinical Microbiology*, **40**: 594–600.

Martens, H., Jensen, S. & Geladi, P. (1983). Multivariate linearity transformation for near-infrared reflectance spectrometry. In: Proceedings of the Nordic symposium on applied statistics. Pp. 205-234. Stokkand Forlag Publishers Stavanger, Norway.

McClure, W.F. (2003) The Giant is running strong. *Journal of Near Infrared Spectrosc*opy, **11**: 487–518.

McDonnel, T. (2019). A fatal public health problem in Africa that flies under the radar: [www.document] URL https://www.npr.org/sections/goatsandsoda/2019/02/21/696385246/a-fatal-public-health-problem-in-africa-that-flies-under-the-radar February, 2019.

Mehl, P.M., Chen, Y, Kim, M.S. & Chan, D.E. (2004). Development of hyperspectral imaging technique for the detection of apple surface defects and contaminations. *Journal of Food Engineering*. Pp. **61**, 67-81.

Mlynáriková, K., Samek, O., Bernatová, S., Růžička, F., Ježek, J., Hároniková, A., Šiler, M., Zemánek, P. & Holá, V. (2015). Influence of Culture Media on Microbial Fingerprints Using Raman Spectroscopy. *Sensors*, **15**, 29635–29647.

Norris K. P. (1959) Infra-red spectroscopy and its application to microbiology. *The Journal of Hygiene* 57(3):326-345.

Nouri, D., Lucas, Y. and Treuillet, S., (2013). Calibration and test of a hyperspectral imaging prototype for intra-operative surgical assistance. *Medical Imaging: Digital Pathology.* Lake Buena Vista, Florida, United States. Pp. 86760.

Park, B., Lawrence, K.C., Windham, W.R. & Smith D.P. (2005). Performance of hyperspectral imaging system for poultry surface faecal contaminant detection. *Journal of Food Engineering*, **75**(3): 340-348.

Park, B., Yoon, S-C., Windham, W. R.,Lawrence, K. C., Kim, M. S., Chao, K. (2011). Line-scan hyperspectral imaging for real-time in-line poultry faecal detection. *Sensory and Instrumentation for Food Quality*, **5**:25–32.

Park, B., Yoon, S. C., Lee, S., Sundaram, J., Windham, W. R., Hinton Jr., A., Lawrence, K. C. (2012). Acousto-optic tunable filter hyperspecral microscope imaging method for characterizing spectra from foodborne pathogens. *American Society of Agricultural and Biological Engineers*, **55**(5): 1997-2006.

Park, B., Seo, Y., Yoon, S.C., Hinton, A., Windham, W.R. & Lawrence, K.C. (2015). Hyperspectral microscope imaging methods to classify Gram-positive and Gram-negative foodborne pathogenic bacteria. *Transactions of the ASABE*, **58**, 5–16.

Peng, Y., Zhang, J., Wang, W., Li, Y., Wu, J., Huang, H., Gao, X. & Jiang, W. (2011). Potential prediction of the microbial spoilage of beef using spatially resolved hyperspectral scattering profiles. *Journal of Food Engineering,* **102,** 163-169.

Qiao, J., Ngadi, M.O., Wang, N., Garie, C. & Prasher S.O. (2007). Pork quality and marbling level assessment using a hyperspectral imaging system. *Journal of Food Engineering* **83**, 10–16.

Qin, J., Chao, K., Kim, M.S., Lu, R. & Burks, T.F. (2013). Hyperspectral and multispectral imaging for evaluating food safety and quality. *Journal of Food Engineering,* **118,** 157-171.

Ramírez-Castillo, F., Loera-Muro, A., Jacques, M., Garneau, P., Avelar-González, F., Harel, J. & Guerrero-Barrera, A. (2015). Waterborne Pathogens: Detection Methods and Challenges. *Pathogens*, **4**, 307–334.

Rinnan, Å., Nørgaard, L., Berg, F.V.D., Thygesen, J., Bro, R. & Engelsen, S.B. (2009a). Chapter 2 - data pre-processing a2 - sun, da-wen. Infrared spectroscopy for food quality analysis and control. Pp.29-50. San Diego: Academic Press.

Rinnan, Å., Van Den Berg, F. & Engelsen, S.B. (2009b). Review of the most common pre-processing techniques for near-infrared spectra. *TRAC Trends in Analytical Chemistry,* **28,** 1201-1222.

Salton, M.R.J., Kim K.S. (2018). Structure. In: Baron S, editor. Medical Microbiology. 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston; 1996. Chapter 2. Available from: https://www.ncbi.nlm.nih.gov/books/NBK8477/

Šašić, S & Ozaki, Y., (2010). Raman, Infrared and Near-infrared chemical imaging. John Wiley & Sons, Inc., Hoboken, New Jersey. Pp. 8 – 20.

Savitzky A, Golay MJ (1964) Smoothing and differentiation of data by simplified least squares procedures. Anal Chem 36(8):1627-1639.

Scott, E., (2003). Food safety and foodborne disease in 21st century homes. *Stanier Review*, **14** (5): 277-280.

Siripatrawan, U., Makino, Y., Kawagoe, Y. & Oshita, S. (2011). Rapid detection of Escherichia coli contamination in packaged fresh spinach using hyperspectral imaging. *Talanta*, **85**(1): p. 276-281.

Suthiluk, P., Saranwong, S., Kawano, S. and Numthuam, S., (2007). Possibility of using near infrared spectroscopy for evaluation of bacterial contamination in shredded cabbage. *International Journal of Food Science and Technology*, **43**,160 – 165.

Tang, A. & Wong, M., (2006). Hazard and risk in food safety: [www document] URL https://www.cfs.gov.hk/english/multimedia/multimedia_pub/multimedia_pub_fsf_01_02.html. August, 2006.

Unusan, N., (2005). Consumer food safety knowledge and practices in the home in Turkey. *Food Control*, **18**: p 45 – 51.

Vargas, A.M., Kim, M.S., Tao, Y., Lefcourt, A.M., Chen, Y.R., Luo, Y., Song, Y. & Buchanan, R. (2005). Detection of Faecal Contamination on Cantaloupes Using Hyperspectral Fluorescence Imagery. *Journal of Food Science*. **70** (8): 471-476.

Vidal, M. & Amigo, J. M., (2012) Pre-processing of hyperspectral images. Essential steps before image analysis. *Chemometrics and Intelligent Laboratory Systems*, **117**: 138-148.

Wang, Y. & Salazar, J.K., (2016). Culture-independent rapid detection methods for bacterial pathogens and toxins in food matrices. *Comprehensive Reviews in Food Science and Food Safety*, **15**: 183-205.

Wang, K., Pu, H., and Da-Wen, S., (2018). Emerging Spectroscopic and Spectral Imaging Techniques for the Rapid Detection of Microorganisms: An Overview. *Comprehensive Reviews in Food Science and Food Safety*, **17**: p. 256-273.

Wilcock, A., Pun, M., Khanona, J. & Aung, M. (2004). Consumer attitudes, knowledge and behaviour: a review of food safety issues. *Trends in Food Science & Technology*, **15**(2): p. 56-66. 26.

Willey, J.M., Sherwood, L.M & Woolverton, C.J., (2011). Prescotts Microbiology. Pp. 54-64.

Williams PJ, Geladi P, Britz TJ, Manley M (2012) Near-infrared (NIR) hyperspectral imaging and multivariate image analysis to study growth characteristics and differences between species and strains of members of the genus Fusarium. *Analytical & Bioanalytical Chemistry* **404**(67): 1759–1769.

Windham, W., Yoon, S-C., Ladely, S.R., Heitschmidt, J.W., Lawrence, K.C., Park, B., Narrang, N., Cray, W.C. (2012). The effect of regions of interest and spectral pre-processing on the detection of non-O157 Shiga-toxin producing *Escherichia coli* serogroups on agar media by hyperspectral imaging. *Journal of Near Infrared Spectroscopy*, **20**(5): p. 547-558.

Yao, H., Hrusk, Z., Kincaid, R., Brown, R.L., & Cleveland, T.E., (2008). Differentiation of toxigenic fungi using hyperspectral imagery. *Sensing & Instrumentation for Food Quality & Safety*, **2**:215–224.

Yoon SC, Lawrence KC, Line JE, Siragusa GR, Feldner PW, Park B, Windham WR (2010). Detection of *Campylobacter* colonies using hyperspectral imaging. *Sensing & Instrumentation for Food Quality & Safety* 4(1):35-49.

Yoon, S.-C., Windham, W.R, Ladely, S.R., Heitschmidt, J.W., Lawrence, K.C., Park, B., Narang, N. and Cray, W.C., (2013a). Hyperspectral imaging for differentiating colonies of non-O157 Shiga-toxin producing

Escherichia coli (STEC) serogroups on spread plates of pure cultures. *Journal of Near Infrared Spectroscopy*, **21**(2): p. 81-95.

Yoon, S.C., Windham, W.R., Ladely, S., Heitschmidt, G.W., Lawrence, K.C., Park, B., Narang, N. & Cray, W.C. (2013b). Differentiation of big-six non-O157 shiga-toxin producing *Escherichia coli* (STEC) on spread plates of mixed cultures using hyperspectral imaging. *Journal of Food Measurement and Characterization,* **7,** 47-59.

CHAPTER 3

Materials and methods

3.1 Sample preparation

The bacterial isolates included in this study were *Bacillus cereus*, *Escherichia coli*, *Salmonella enteritidis*, *Staphylococcus aureus* and *Staphylococcus epidermidis*. Standard bacterial strains from the American Type Culture Collection (ATCC) were sourced from ThermoFisher Scientific laboratories, South Africa. The ATCC numbers are as follow: *Bacillus cereus* (ATCC 13061), *Escherichia coli* (ATCC 25922), *Salmonella enteritidis* (ATCC 13076), *Staphylococcus aureus* (ATCC 29213) and *Staphylococcus epidermidis* (ATCC 12228). A second isolate of *S. aureus* (ATCC 25923) was obtained from an established culture collection at the Department of Food Science, Stellenbosch University.

To investigate the effect of growth media on the spectra, two different general-purpose growth media were selected: nutrient agar (NA) and tryptic soy agar (TSA) (Neal, 2019). Both a streak plate and spread plate method was also employed as different plating techniques could potentially influence the spectra.

An initial stock solution was made by suspending 0.1 mL frozen stock culture into 5 mL Tryptic soy broth (TSB) and was incubated overnight at 37 °C for 24 h. The overnight suspension contained approximately 10^8 colony-forming unit (CFU) mL^{-1} bacterial cells. For the streak plate method, each bacterium was streaked onto solid NA or TSA growth medium in 80 x 15 mm glass Petri dishes (Kammies, 2018). For the spread plate method, a seven-fold (10^8 to 10^1 (CFU) mL^{-1}) serial dilution was prepared by adding 1 mL stock solution to 9 mL Tryptic soy broth (TSB). Exactly 0.1 mL aliquots of the 10^1 (CFU) mL^{-1} dilution was then inoculated onto NA and TSA agar plates separately, using a sterile glass hockey-stick to spread it evenly over the surface of the agar. The procedure was carried out under aseptic conditions in a biosafety fume hood and all samples were prepared in triplicate.

Following the above protocols, separately, over two days, the Petri-dishes were then incubated for 24 h at 37 °C after each batch was complete. Both streak plates and spread plates yielded single colonies (Wiley *et al*., 2011), on both growth media used (Figure 3.1). Before image acquisition, the plates were allowed to cool down to ambient temperature (approximately 23 °C).

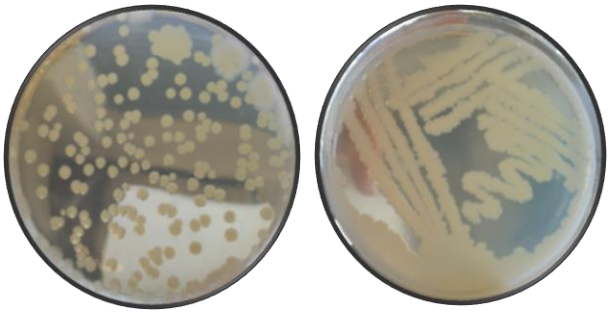

Figure 3.1 Digital images (obtained with HUAWEI P20 12 MP (RGB, f/1.8 aperture) + 20 MP (Monochrome, f/1.6 aperture)) of (left), *S. enteritidis* on NA agar using spread plate method, (right), *S. epidermidis* on NA using a streak plate method.3.2 NIR-HSI Imaging system.

3.2 NIR – HSI Imaging system

Images were acquired with the SWIR-384 (short wave infrared) push-broom hyperspectral imaging system (Norsk Elektro Optikk, Norway), operated with Breeze® (Prediktera AB, Umeå, Sweden) software version 2019.2.0. The samples were placed onto a translation stage which moved at a speed of 50 mm/s and the camera, which was equipped with a 30 cm focal length lens, was mounted onto a laboratory rack. The spectral range used was 950-2500 nm with 384 spatial pixels and 288 spectral channels. A spatial resolution of 0.247 mm pixel size and 5.45 nm spectral sampling per pixel was used. The SWIR camera of the system provided a field of view (FOV) of 9.47 cm and a dynamic range of 7500. The system was equipped with two 150 W halogen lamps (Ushio lighting Inc., Japan) that has the capacity to emit light in the 400 – 2500 nm wavelength range. The lamps were symmetrically situated at a distance of 30 cm above the translation stage. Images were acquired at a maximum frame rate of 100 frames per second (fps) with a manually set integration time of 3200 µs.

3.2.1 Image acquisition

Prior to image collection, a radiometric calibration was automatically done in the Evince software package using **eq. 3.1** (Sendin, 2017). (See section 3.5, **eq. 2.2** for corrected reflectance value). The white standard used was a 50% Zenith Allucore grey diffuse reflectance standard (SphereOptics GmbH, Germany) and was used in combination with recorded dark references to correct for uneven light intensities.

$$l_{\lambda,n} = -log_{10}\left[\left(\frac{S_{\lambda,n}-\mathrm{B}_{\lambda,n}}{W_{\lambda,n}-B_{\lambda,n}}\right)\right] \quad \text{(eq. 3.1)}$$

Where:

n = Pixel index variable (n = 1…N) of the reorganised hypercube

$I\lambda$, = Standardised absorbance intensity, pixel n, at wavelength λ

$S\lambda$, = Sample image, pixel n, at wavelength λ

$B\lambda$, = Dark reference image, pixel n, at wavelength λ

$W\lambda$, = White reference image, pixel n, at wavelength λ

All Petri-dishes were allowed to cool down to ambient temperature for approximately 20 min (Kammies *et al*., 2016). Each Petri-dish (as prepared in triplicate), containing bacteria on either NA or TSA, was then imaged, with the lids removed. The samples were then discarded accordingly.

3.3 Data Analysis

3.3.1 Principal Component Analysis (PCA)

A PCA model was calculated for individual images to remove unwanted background pixels (Esbensen *et al.,* 2006) It was important to isolate only the bacterial growth by removing the agar background and redundant pixels. Since glass Petri-dishes were used, some of the samples contained shadows and therefore also had to be removed. The score plots generated individual clusters that represents, bacteria, agar and background information. This could be highlighted and removed or retained where necessary. This was an iterative process, repeated until all unwanted information was removed. Once all the images were cleaned, they were merged to form mosaics. The wavelengths from 2255.5 nm – 2500 nm range were removed as they contained noisy areas and were deemed unnecessary for further analysis. Different combinations of mosaics containing different bacteria were grouped together depending on the objective.

3.3.2 Mosaic Construction

Six foodborne pathogenic isolates were investigated using a two-part objective. The first, was to evaluate the effect of different growth media on the NIR spectra of bacteria. Four groups of mosaics were examined independently for this objective. The plating technique was kept constant and therefore only the streak plates were evaluated. Both the pixel wise and object wise approaches were investigated to determine which provided the best results (Kucheryavskiy, 2013).

Pixel wise analysis vs Object wise analysis

With the pixel wise approach, each pixel in the image provides a spectrum which is used in the PCA calculation. In contrast to this, in the object wise approach, an average spectrum of each object (which is selected manually) is calculated and then used in the PCA calculation. The manual selection of the objects was done by segmenting each individual sample into approximately 25 separate 'objects'. A schematic of the process is illustrated in Figure 3.2 where streak plates of the Gram-negative bacteria were used as an example.

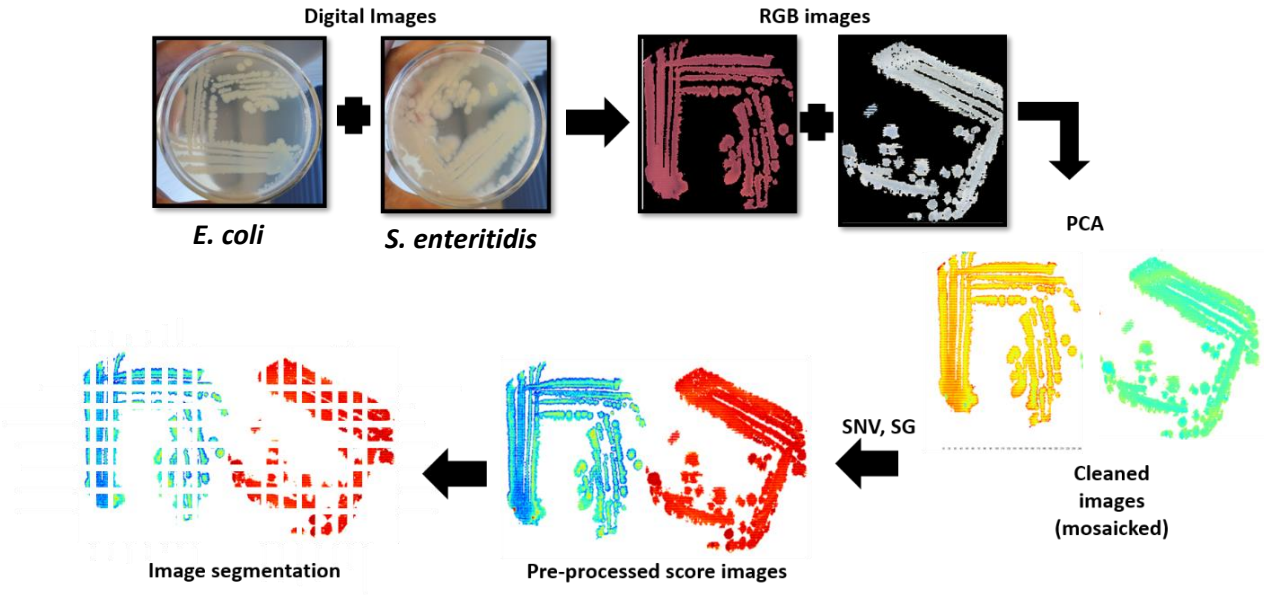

Figure 3.2 Schematic of mosaic construction, image processing and image segmentation sequence followed in this study.

First, we evaluated the pixel-wise approach for the differences between each bacterial isolate on different growth media separately. **Group 1** contained all the isolates on NA and **Group 2** contained all the isolates on TSA. These two mosaics were constructed to evaluate if the organisms could be separated on the different growth media (NA & TSA) separately. An object-wise analysis was then carried out for both to determine which approach would give optimal results. The analysis (pixel-wise and object-wise) continued in a similar manner for **Group 3** however, we evaluated the effect of the growth media on the separation of the bacteria based on their Gram-stain classification. This group therefore contained all the isolates on NA and TSA, now further grouped together as either Gram-positive or Gram-negative.

The final analysis was done to distinguish between pathogenic and non-pathogenic isolates. **Group 4** therefore contained the two *S. aureus* isolates (referred to this as *S. aureus*1 (ATCC 29213) and *S. aureus*2 (ATCC 25923) respectively, in the rest of this thesis) and *S. epidermidis* using an object-wise approach.

For the second objective, seven mosaics were constructed to evaluate the effect of different plating techniques on the separation of bacteria based on their Gram-stain classification. **Group 5** and **6** comprised of Gram-positive bacteria using a streak and spread plate technique, respectively. **Group 7** contained Gram-positive bacteria using a combination of the two techniques. Similarly, **Group 8** and **9** contained Gram-negative bacteria using a streak and spread plate technique respectively. **Group 10** contained Gram-negative bacteria using the two techniques combined. Finally, **Group 11** comprised of all six isolates using both plating techniques.

PCA was recalculated and two pre-processing techniques, standard normal variate (SNV) and Savitzky-Golay smoothing filter (2nd derivative, 3rd order polynomial, 25-point smoothing) were applied for **Groups 1 - 4.** However, to further optimise the separation of the bacteria, three different pre-processing techniques ((SNV, SGd_2; (9pt.) and SGd_2; (5pt.)) were applied to **Groups 5-11** to determine which would obtain better results**.** To observe any differences between the bacteria, score images, score plots and loading line plots were analysed interactively. To select the regions of interest, the score images and score plots along with the digital images were used interactively. The digital images were used to confirm the data obtained from the abovementioned plots. Additionally, the mean spectra for each bacterium was computed and analysed.

3.3.3 Partial Least Squares Discriminant Analysis (PLS-DA)

A PLS-DA model was then calculated for each group. Both SNV and Savitzky-Golay (2nd derivative, 3rd polynomial, 25-point smoothing) pre-processing was applied to all the models. The PLS-DA algorithm (**eq. 3.2**) aims to separate the data into groups by finding a straight line that divides a given space into two regions (Chavallier *et al*., 2006; Amigo *et al*., 2013).

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{f} \qquad \textbf{(eq. 3.2)}$$

Where:

y = Prediction of indicator variables

X = Unfolded hypercube obtained from hyperspectral image

β = Beta coefficient

f = Residual or 'error' matrix

Matlab (2017b), was used to perform the Venetian blinds cross-validation method (Eigenvector, 2016; PLS Toolbox, 2016), on the object wise PLS-DA calibration model. The pixel wise models contained a far too large data set for this computation. The calibration model was used to determine the optimum number of latent variables (LV's). The PLS-DA model computes the coefficient of determination (R^2), closest to one and is used as

an indication of the optimal number of LV's. The calibration set contained 70% of the total samples and the remaining 30% was the test set. This split was computed using the Duplex data splitting method, which uses the Euclidean distance between the two objects furthest from each other to create the initial calibration set. The two successive objects further from each other is used to create the initial test set. The process is repeated until all the samples have been assigned to a specific set (Snee, 1977). The models were independently validated using the optimal conditions (number of LV's) on the portion of the total samples represented in the test set to obtain the prediction results.

A confusion matrix was then obtained containing the classification accuracy percentage of the model. From the matrix, the classification accuracy, which illustrates the overall effectiveness of the model, was obtained using equation 3.3. The false positive error and false negative error were also obtained using equation(s) 3.4 and 3.5 respectively. The sensitivity of the model, also referred to as the true positive rate, was calculated using equation 3.6 and refers to the probability that a given organism in the model would be correctly classified. This gives an indication of the true effectiveness of the model based on different classes. The specificity is indicative of the probability of the model to correctly predict the remaining organisms using equation 3.7 and is also referred to as the true negative rate. The classification accuracy is simply the rate of correctly classified objects in a sample and can be computed using the number of true positives and true negatives. This can be confirmed by calculating the sensitivity and specificity. The predictive power of the model is indicated when calculating the precision. This is done by estimating the value predicted for each class, as per equation 3.8. Lastly, the misclassification rates for each model can be calculated using equation 3.9. Calculations for the results obtained from the confusion matrix are as follows:

$$\text{Classification accuracy (\%)} = \frac{True\ positives + True\ Negatives}{Total}\ x\ 100\% \quad \text{(eq. 3.3)}$$

$$\text{False positive error (\%)} = \frac{False\ positives}{Total}\ x\ 100\% \quad \text{(eq. 3.4)}$$

$$\text{False negative error (\%)} = \frac{False\ Negatives}{Total}\ x\ 100\% \quad \text{(eq. 3.5)}$$

$$\text{Sensitivity or Recall (\%)} = \frac{True\ positives}{True\ positives - False\ negatives}\ x\ 100\% \quad \text{(eq. 3.6)}$$

$$\text{Specificity (\%)} = \frac{True\ Negatives}{True\ Negatives + False\ positives}\ x\ 100\% \quad \text{(eq. 3.7)}$$

$$\text{Precision (\%)} = \frac{True\ Positives}{True\ Positives + False\ positives}\ x\ 100\% \quad \text{(eq. 3.8)}$$

$$\text{Misclassification rate (\%)} = \frac{False\ Negatives + False\ Positives}{Total}\ x\ 100\% \quad \text{(eq. 3.9)}$$

Where:

(using a model to distinguish Gram-positive bacteria from Gram-negative bacteria as an example)

True Positives = Gram-positive bacteria classified correctly

True Negatives = Gram-negative bacteria classified correctly

False positive = Gram-negative bacteria, incorrectly classified as Gram-positive bacteria

False negative = Gram-positive bacteria, incorrectly classified as Gram-negative bacteria OR not classified at all

Total = the sum of all the organisms used in the given model

The remaining images were used to validate the models. The prediction map is used as a visual representation of the overall performance of the models. The sum of the organisms was used to determine the number of correctly and incorrectly classified samples, using the object wise approach. Pixels that were unassigned in the model were predicted as “no class” and regarded as false negatives for this study.

3.4 REFERENCES

Amigo JM, Martí I, Gowen A (2013) Hyperspectral imaging and chemometrics: A perfect combination for the analysis of food structure, composition and quality. Marini F (ed) Chemometrics in Food Chemistry. 1st edn. Elsevier, pp 343-370

Chevallier S, Bertrand D, Kohler A, Courcoux P (2006) Application of PLS-DA in multivariate image analysis. J Chemometrics 20(5):221-229.

Eigenvector Research, (2016). Venetian Blinds, *Using Cross-Validation*: [www document] URL http://www.eigenvector.com/index.php?title=Faq_reference_Eigenvector

Esbensen, K.H., Guyot, D., Westad, F. & Houmolle, L.P. (2006). An Introduction to Multivariate Data Analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **169**, 387–389.

Kammies, T. (2018). The evaluation of foodborne pathogenic bacteria using near infrared (NIR) hyperspectral imaging and multivariate image analysis. *MSc Thesis*. Stellenbosch University, Stellenbosch.

Kammies, T.L., Manley, M., Gouws, P.A. & Williams, P.J. (2016). Differentiation of foodborne bacteria using NIR hyperspectral imaging and multivariate data analysis. *Applied Microbiology and Biotechnology*, **100**, 9305–9320.

Kucheryavskiy, S. (2013). A new approach for discrimination of objects on hyperspectral images. *Chemometrics and Intelligent Laboratory Systems*, **120**, 126–135.

MATLAB and Statistics Toolbox Release 2017b*, The MathWorks*, Inc., Natick, Massachusetts, United States

Neal, E., (2019). Types of agar plates. *Sciencing*: [www document] URL https://sciencing.com/types-agar-plates-8131230.html. 11 August 2019.

PLS_Toolbox with MIA_Toolbox 8.2.1 (2016). Eigenvector Research, Inc., Manson, WA USA 98831; software available at http://www.eigenvector.com.

Sendin, K., (2017). Characterisation of whole white maize kernels. *MSc Thesis*. Stellenbosch University, Stellenbosch.

CHAPTER 4

Evaluating NIR-HSI as a complimentary rapid screening tool for food microbiology.

4.1 Effect of different growth media on the NIR spectra of bacteria

4.1.1 Spectral Analysis

The average spectra obtained for the Gram-stain reaction model on different growth media is depicted in the image below (Figure 4.1).

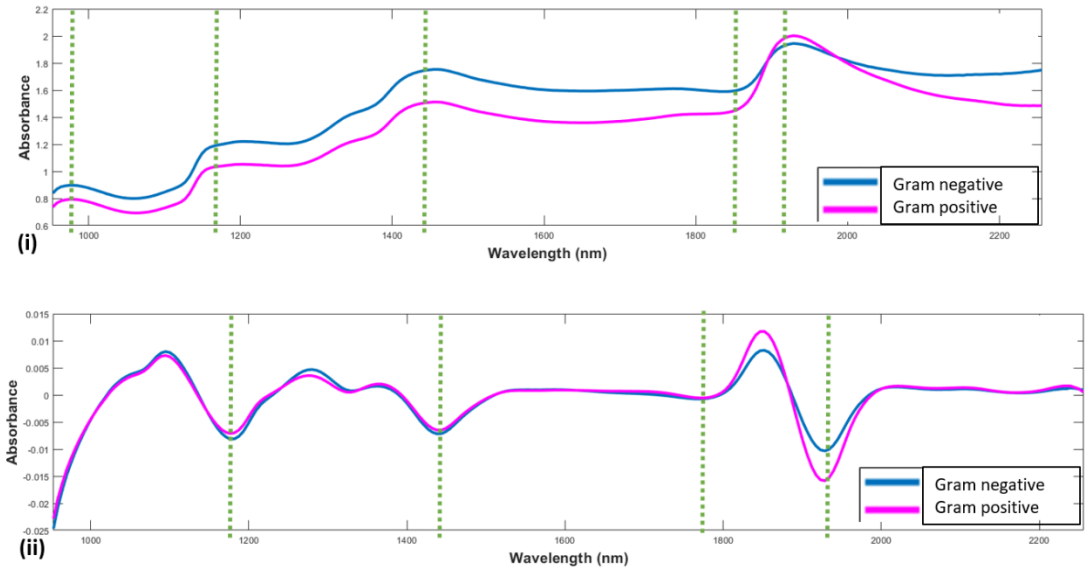

Figure 4.1 (iii) Average spectra, unprocessed (in the spectral range 950 – 2255 nm) for Gram-positive isolates: *Staphylococcus aureus* (ATCC 29213 & ATCC 25923), *Staphylococcus epidermidis* and *Bacillus cereus*. Gram-negative isolates: *Escherichia coli* and *Salmonella enteritidis*. (ii) Pre-processed spectra: the data was pre-processed with Standard normal variate and Savitzky-Golay (2nd derivative, 3rd polynomial, 25pt. smoothing filter) in both pixel-wise and object-wise analysis.

The overall spectral results show very similar absorption bands for the different bacteria, with a slight difference in intensity in certain regions. The energy absorption in the NIR region is as a result of overtones and combinations of fundamental vibrations of certain functional groups. The absorption bands in the NIR region tend to be quite broad and could potentially hide peak dissimilarities, the derivative spectra therefore makes visualising the peak differences more apparent (Maquelin *et al.*, 2000). The prominent bands at 975 and 1454 nm are related to O-H stretching in the first and second overtone respectively, attributed to water (Osborne, 2000; Dubois *et al.*, 2005; Kammies *et al.*, 2016). Whereas the peak near 1928 nm is related to an O-H bend in

the second overtone and deformation combination, as well as the N-H stretch in the second overtone which could be related to the CONH group observed in proteins (Wilson *et al.*, 2015; Bezuidenhout, 2018). Water is strongly absorbed in the NIR region and makes up about 70% of the composition of bacteria (Cooper *et al*., 2000; Dubois *et al.*, 2005). The water absorption features are therefore as expected. The most pronounced bands reported for moisture are at 1450 and 1950 nm (Curran *et al.*, 1997; Clevers and Kooistra, 2006). The band at 970 nm is not reported as often but is still observed. In the spectral bands related to moisture, atmospheric absorption of water vapour is strong (Clevers and Kooistra, 2006). A minimal amount of water vapour was observed around the edges of the plates and could therefore be the cause for the deviations from the fundamental absorption bands.

The lipid content in the sample is represented by the absorption band at 1180 nm, which is related to the C-H stretch in the second overtone, observed in the amide III region (Osborne, 2000; Wilson *et al.*, 2015; Mosier-Boss, 2017). Similarly, the peak at 1792 nm is related to C-H stretching in the first overtone and was also attributed to lipid content (Osborne, 2000; Rodriguez-Saona *et al.*, 2001). Different wavelengths are associated with different molecular vibrations and so the spectral absorption bands highlighted here cannot solely be as a result of the chemical composition of foodborne pathogens (Nakakimura *et al.*, 2012). The raw spectra may contain light scattering effects and the spectra could be a mixture of the organism and the agar (Feng and Sun, 2014). Savitzky-Golay smoothing filter, resolves broad and overlapping peaks making it easier to observe differences in the spectra. Examining the pre-treated, individual spectra is therefore essential to determine which bands are responsible for the differences in the spectra (spectral results of individual bacteria on NA and TSA in comparison to Figure 4.1 can be seen in **Addendum A** (FigureA1).

Data Analysis

4.1.2 Principal component analysis (PCA)

Group 1: PCA results of bacteria – NA

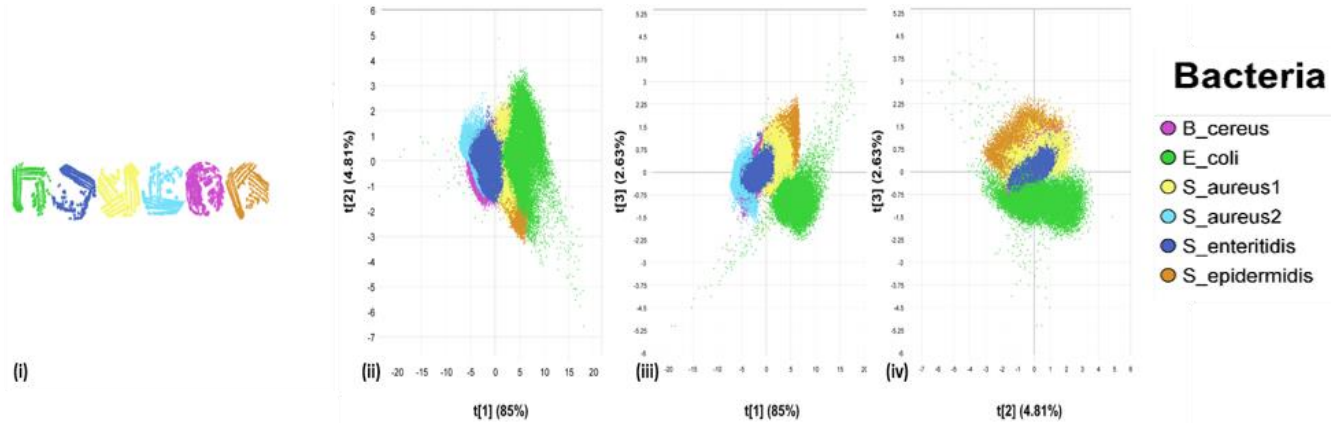

Figure 4.2 Pixel-wise PCA of Group 1 illustrates the individual bacteria on Nutrient agar (NA). PCA score image (i), shaded according to the different bacterial isolates (green: E. coli, blue: S. enteritidis, purple: B. cereus, yellow: S. aureus1, light blue: S. aureus2 and orange: S. epidermidis).

Figure 4.3 Objects created by segmentation of bacterial samples into approximately 18 objects per sample, obtaining a total of 107 objects.

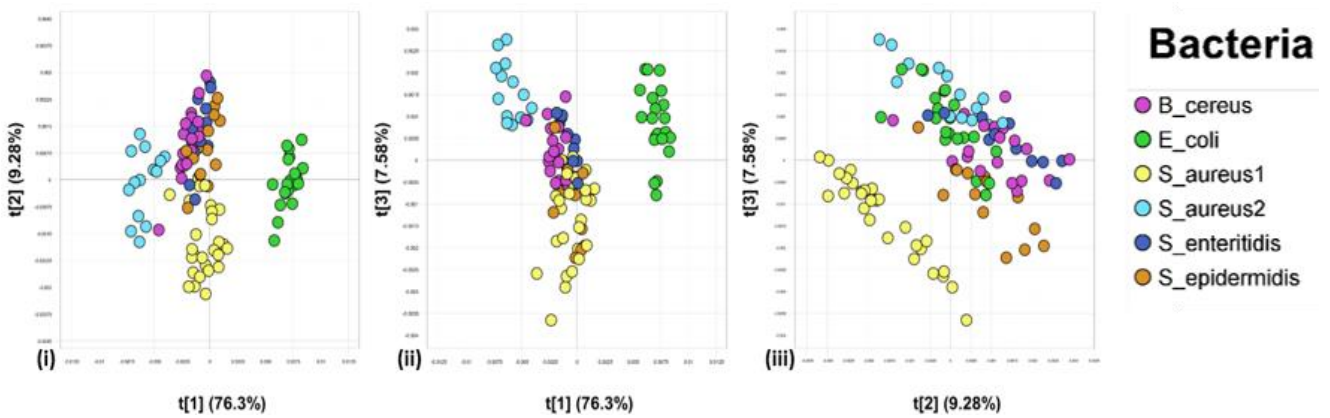

Figure 4.4 Object-wise PCA of Group 1 illustrating improved separation between the bacterial isolates. The isolates were coloured per bacteria as seen in the legend. The PCA score plot (i), of PC1 (76.3%) vs. PC2 (9.28%) the E. coli isolates clustered together, confirming that separation is possible. In PCA score plot (ii), of PC1 (76.3%) vs. PC3 (7.58%), a similar observation was made. In PCA score plot (iii), PC2 (9.28%) vs. PC3 (7.58%), the S. aureus1 samples clustered together separating it from the rest of the isolates.

Group 2: PCA results of bacteria – TSA

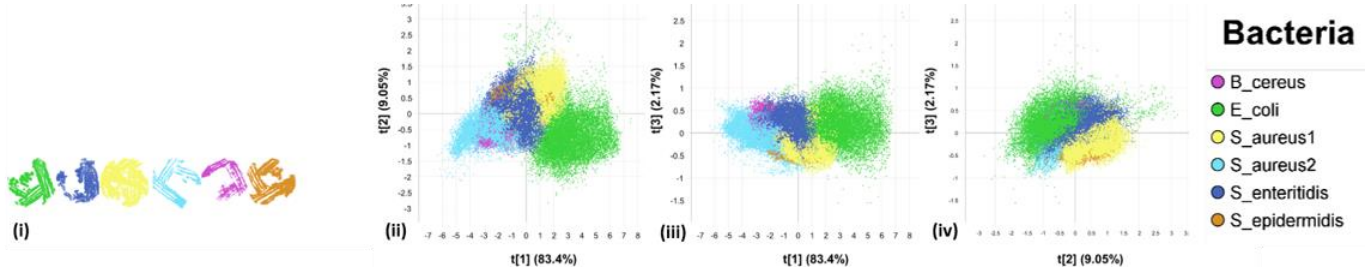

Figure 4.5 Pixel-wise PCA of Group 2 illustrates the individual bacteria on Tryptic soy agar (TSA). In the PCA score image (i), the different bacterial isolates can be seen (LTR: E. coli, S. enteritidis, B. cereus, S. aureus1 (ATCC 29213), S. aureus2 (ATCC 25923) and S. epidermidis).

Figure 4.6 Objects created by segmentation of bacterial samples into approximately 23 objects per sample, total objects of 137 objects.

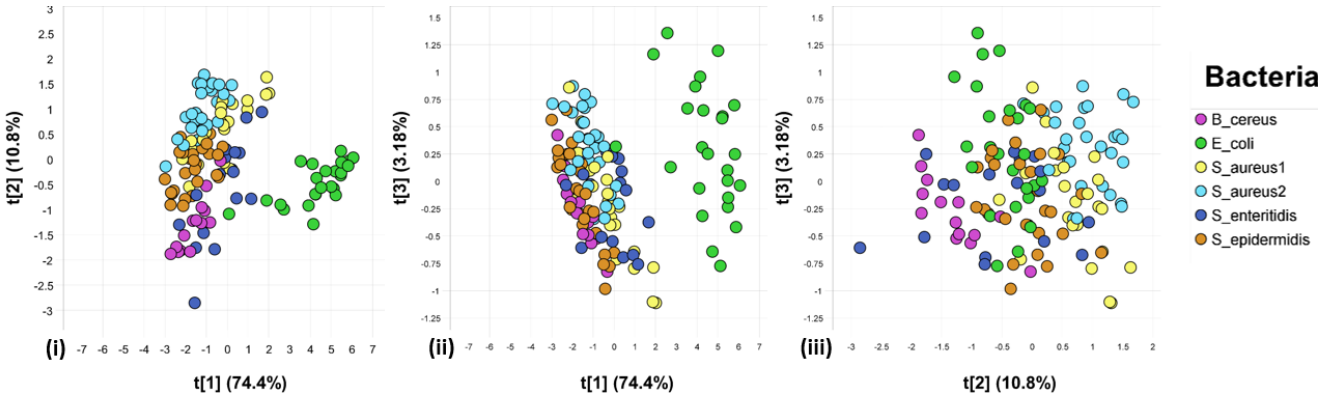

Figure 4.7 Object-wise analysis (Group 2), showing improved separation between the bacterial isolates on TSA. The PCA score plot (i), of PC1 (74.4%) vs. PC2 (10.8%). In PCA score plot (ii), in the direction of PC1 (74.4%) vs. PC3 (3.18%), a similar observation was made. In PCA score plot (iii), PC2 (10.8%) vs. PC3 (3.18%), the samples were scattered randomly – no clear distinction.

Comparing Group 1 & 2 Pixel-wise and Object-wise analysis

When comparing the pixel-wise results of both **Group 1** & **2**, clustering of the individual observations is shown. In Figure 4.2, PCA score plot (i) of PC1 (85%) vs. PC2 (4.81%) compared to Figure 4.5, PCA score plot (i) of PC1 (83.4%) vs. PC2 (9.05%), no clear separation of the individual bacteria is observed. There is a slight decrease (1.6%) in overall variance accounted for by PC1 in **Group 2**. However, in PC2 an increase of 4.24% is seen for the total variance in the sample. In PC1 (85%) vs. PC3 (2.63%) (Figure 4.2 PCA score plot (ii)), in comparison to PC1 (83.4%) vs. PC3 (2.17%) (Figure 4.5 PCA score plot (ii)), we see a small decrease of 0.46% accounted for by PC3. Lastly, when comparing PC2 (2.63%) vs. PC3 (4.81%) (Figure 4.2 PCA score plot (iii)), to PC2 (9.05%) vs. PC3

(2.17%) (Figure 4.5 PCA score plot (iii)), again no separate clustering of the individual bacteria is observed. In both **Group 1** (NA) and **2** (TSA), PC1 is responsible for the majority of the variance observed in the sample. The pixel-wise score plots suggest that the *E*. *coli* isolates are strongly associated with the positive region in the PCA score plots, however this would have to be confirmed by examining the loadings (Figure 4.8) and mean spectra (**Addendum A**, Figure A3) of the individual bacteria.

When comparing the object-wise results of score plot of PC1 (76.3%) vs. PC2 (9.28%) (Figure 4.4(i)) to that of score plot of PC1 (74.4%) vs. PC2 (10.8%) (Figure 4.7(i)), we see a slight decrease of the percentage variance accounted for by PC1. However, in **Group 2**(PC1), the samples are packed quite densely, showing clustering of similar attributes. In PC2 an increase of 1.52% is observed. It is unclear whether this is on account of the agar or the cell wall components of the individual bacteria. For this reason, the average spectra need to be examined. (**Addendum A,** FigureA2 & A3). In PC1 (76.3%) vs. PC3 (7.58%), Figure 4.4(ii) compared to PC1 (74.4%) vs. PC3 (3.18%), Figure 4.7(ii), we observe that an increase of 4.4% in the total variance accounted for PC3. The Gram-negative, *E*. *coli* samples, forms a clear cluster in the positive region of the PCA score plots in both groups, whereas the *S*. *aureus* samples, seem to be strongly associated with the negative region. The *E*. *coli* samples were more densely packed in **Group 1** (NA) Figure 4.4(ii), compared to the samples in **Group 2** (TSA) Figure 4.7(ii) where the isolates were scattered randomly and are more spaced out. Finally, when comparing PCA score plot of PC2 (9.28%) vs. PC3 (7.58%) Figure 4.4(iii), to that of PCA score plot of PC2 (10.8%) vs. PC3 (3.18%) Figure 4.7(iii), the *S*. *aureus* samples in **Group 1** (**NA**) forms a cluster in the positive region of the PCA score plot separating it from the rest of the isolates. However, a random arrangement of the samples is seen for **Group 2**, showing no clear distinction for the TSA samples.

The overall pixel-wise results were not encouraging as the bacterial isolates could not be clearly separated on the different growth media. The object-wise approach however, showed improved separation by means of cluster formation for some of the samples. In score plot of PC1 (85%) vs. PC2 (4.81%), Figure 4.2(i), which accounted for the majority of the total variance in the sample (pixel-wise), compared to the (object-wise) score plot of PC1 (76.3%) vs. PC2 (9.28%), Figure 4.4(i), although a decrease of 8.7% is observed, the separation of the individual isolates improved. The same is true for score plot of PC1 (83.4%) Figure 4.5(i), compared to the score plot of PC1 (74.4%) Figure 4.7(i), where a decrease of 9% in the total variance can be seen. However, in the object-wise analysis, the mean spectra for each individual object are used in the calculation of the PCA, simplifying the task of finding similar attributes in the samples to a certain extent (Kucheryavskiy, 2013). The overall models for the NA and TSA samples showed very similar results. Although it is unclear whether the separation observed is as a result of the different media or the biochemical characteristics of each individual isolate, the results show that irrespective of the growth media used, using an object-wise approach distinction between different foodborne pathogenic bacteria is possible.

The loading plots were used interactively with the PCA score plots to better understand these results. Important wavelengths contributing to the variance in the data is found in the loading lines. Since the same data was used for both the pixel-wise and object-wise approach, the loadings observed were similar. The observations made in the negative region of the loading plots correspond to the negative region in the corresponding PCA score plot and vice versa (Esbensen *et al.*, 2006).

Loadings Group 1

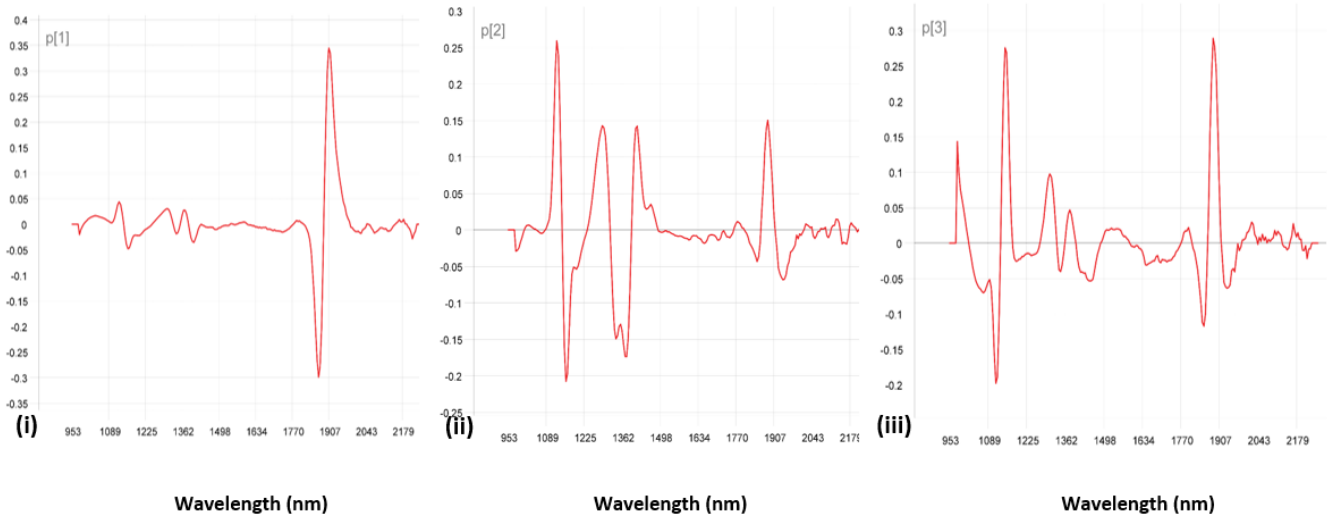

Figure 4.8 PCA loading lines of **Group 1** (all six isolates) on NA. Important peaks were observed for (i) PC1 at 1907 nm (pos. region) and at 1868 nm (neg. region). For (ii) PC2 at 1127, 1291, 1416 and 1885 nm (pos. region) and in the negative region at 1160 and 1378 nm. The prominent peaks in (iii) PC3 (pos. region) are observed at 980, 1149, 1307, 1378 and 1885 nm and in the negative region at 1116, 1345 and 1852 nm.

When closely examining the loadings of **Group 1** (Figure 4.8), along with the object-wise PCA score plots of the same group (Figure 4.4), the prominent peak observed in loading 1 (i), shows the separation of *S*. *aureus*1 from the rest of the isolates which is mostly based on the band at 1868 nm in the negative region. Some of the *S*. *aureus*2 isolates are also seen in this region, which shows that there are inter-strain similarities. The band is related to the O-H and C=O stretch in the first overtone and combination band region which is indicative of a carbohydrate (Osborne, 2000; Genot *et al.*, 2014). The positively loaded band at 1907 nm is associated with the *E*. *coli* isolate which is clearly separated from the rest of the isolates. This observation is also evident in the score plot of **Group 1** (Figure 4.4). This peak is related to the O-H and NH_2 stretch in the first overtone and combination band region (Genot *et al.*, 2014; Wilson *et al.*, 2015). The separation could be ascribed to the water-binding capacity of *E*. *coli* in the NIR region or could be as a result of a unique protein structure found the cell wall that separates it from the rest (Dubois *et al.*, 2005).

In loading 2 (Figure 4.8(ii)), in the positive region, we observed bands at 1127 and 1291 nm as a result of lipid content in the sample. These bands are related to the C-H stretch in the second overtone and were associated with *S*. *aureus*1, *B*. *cereus*, as well as some of the *Salmonella* isolates which shows that there are underlying similarities within the different Gram stain reaction (Gram-positive and Gram-negative) groups. The band at 1416 nm, was associated with *S*. *aureus*1 and *B*. *cereus* and could be as a result of protein content. This correlates with the O-H and N-H stretch in the first overtone (CONH) (Osborne, 2000; Wilson *et al.*, 2015). In the negative region the bands at 1160 and 1378 nm is associated with the *S*. *aureus*1 strain which was attributed to lipids.

When examining (Figure 4.8(iii)), loading 3 compared to PCA score plot of PC1 (74.4%) vs. PC3 (3.18%) (Figure 4.4(ii)), the *S*. *aureus*2 sample is responsible for the strong absorption of water at 980 nm in the positive region. The peak at 1149 nm was attributed to carbohydrates in *S*. *aureus*1, whereas both *S*. *aureus* samples and some of the *E*. *coli* samples (data points) are responsible for the peaks at 1307 and 1378 nm as a result of lipid content. Again, showing similar properties within the different Gram-stain reaction isolates. An intense peak is observed at 1885 nm due to the presence of *E*. *coli*, which is as a result of carbohydrate content. The PCA score plot of PC2 vs. PC3 (Figure 4.4(iii)), shows clear separation of *S*. *aureus*1 from all the other isolates, associated with the negatively loaded bands at 1116 and 1345 nm, attributed to carbohydrate and lipid content respectively. Lastly, the band at 1852 nm is strongly associated with the presence of *S*. *aureus*1 in the sample as a result of carbohydrate content.

Loadings Group 2

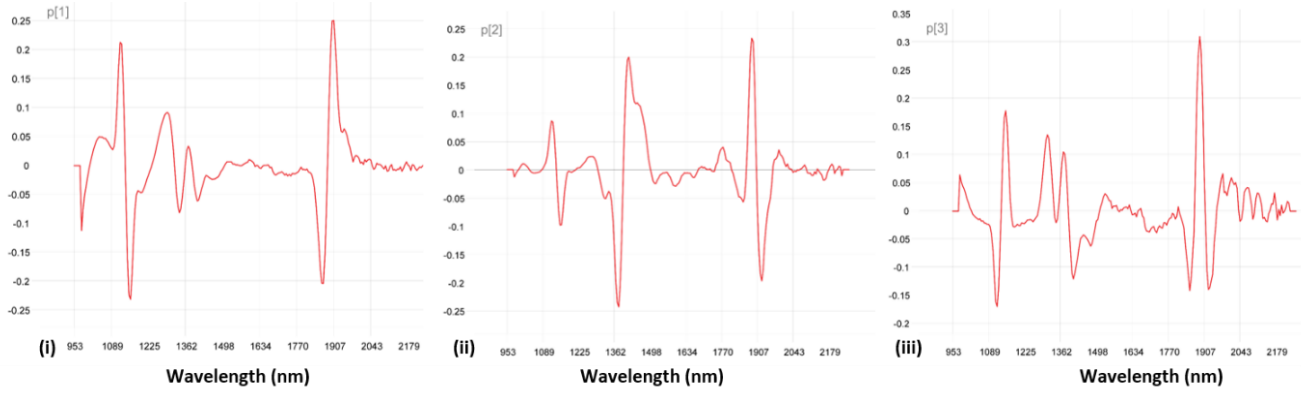

Figure 4.9 PCA loading lines of **Group 2** (all six isolates) on TSA. The important bands contributing to the variance in the sample in (i) PC1 were observed at 1122 nm and 1907 (pos. region) and at 980, 1160, 1340, 1372 and 1885 nm (neg. region). For (ii) PC2 at 1122, 1416 and 1885nm (pos. region) and at 1160, 1378 and 1923 nm in the

negative region. The prominent peaks in (iii) PC3 (pos. region) are observed at 980, 1154, 1313, 1345 and 1890 nm and lastly, in the negative region, we observe prominent peaks at 1122, 1378, 1411, 1852 and 1923 nm.

The important peaks in loading 1 (Figure 4.9) for **Group 2** were observed in (i) at 1122 and 1907 nm in the positive region. The peak at 1122 nm in the positive region is associated with the lipid content in the sample. When examining the PCA score plot of PC1 vs. PC2 (Figure 4.7(i)), we observe that this is where *E*. *coli* is located. Some of the *Salmonella* samples are also located here, confirming the Gram-stain similarities of these two isolates. A prominent peak is seen at 1907 nm which is as a result of the O-H bend in the second overtone and deformation combination region, however could also be as a result of protein content associated with the N-H stretch in the second overtone as stated by Dubois *et al.*, (2005). This positively loaded peak is strongly associated with the *E*. *coli* samples, which is clearly separated from the rest of the isolates (Figure 4.7(i)). The intense peak at 980 nm in the negative region is due to the water absorption of the organisms. The *B*. *cereus* sample is strongly associated with this band as it shows the second highest intensity and the separation can be seen in PCA score plot of PC1 (74.4%) vs. PC2 (10.8%) (Figure 4.7 (i)). The bands at 1340 and 1372 nm show separation of *B*. *cereus* and some of the *Salmonella* samples that are also located in PCA score plot of PC1 vs. PC2 (Figure 4.7(i)) due to lipid content. Again, showing similar properties within the different Gram-stain reaction isolates. An intense peak is observed at 1885 nm, due to the presence of these two isolates. When looking at the digital images (**ADDENDUM A**, Figure A6), we observed that these two isolates suffered from overgrowth as a result of the 24-hour incubation period. This suggests that there could be some overlapping between these two isolates due the thickness of the colonies. The growth densities could be attributed to the composition of the agar and could thus be the reason for the lack of separation of the two organisms.

In loading plot 2 (Figure 4.9(ii)), we observe a peak in the positive region at 1122 nm, due to the presence of lipid content likely due to the presence of the *S*. *aureus* isolates. The Gram-positive cell wall is a macromolecule composed of peptidoglycan that consists of sugars and amino acids (Navarre and Schneewind, 1999). A vast amount of cell surface molecules such as teichoic acids, teichuronic acids, polyphosphates and carbohydrates are present in the Gram-positive cell wall and not present in the Gram-negative cell wall. The proteins observed in the periplasm of the Gram-negative cell wall are lipid-modified in the Gram-positive cell wall and could therefore be a reason for this observation. When examining the PCA score plot in the direction of PC2 (Figure 4.7(i)), we observe that the presence of the *S. aureus* is predominantly responsible for the peak at 1416 nm (protein) and the intense peak at 1885 nm (carbohydrate). The *S. aureus* and *B. cereus* (Gram-positive) isolates are present in the negative region of PC2 (Figure 4.9(ii)), at 1160 nm (carbohydrate). The *B. cereus* isolate is solely responsible for the intense peaks at 1378 and 1923 nm, attributed to carbohydrate and lipid content respectively. This correlates with the PCA score plot of PC1 vs. PC2, showing the separation of the *B. cereus* isolate in the negative region.

Lastly, when examining the associated PCA score plot along with that of loading 3 (Figure 4.9(iii)), we observed that the Gram-negative isolate (*E*. *coli*) is responsible for the positively loaded prominent peaks, whereas all the Gram-positive isolates were predominantly responsible for the negatively loaded prominent peaks. The band at 980 nm was due to the presence of *E*. *coli*. The band at 1154 nm is associated with the *S*. *aureus* isolates in the sample, due to similar lipid content. In the negative region, we see an intense peak at 1122 nm also due to the lipid content of *S*. *aureus*1. A similar observation is made at 1372 and 1411 nm due to the respective carbohydrate, lipid and protein content present in both *S*. *aureus*1 and *S*. *epidermidis*. Lastly, the band at 1852 nm is associated with the presence of *Salmonella* in the sample as a result of carbohydrate content, whereas the negatively loaded peak at 1923 nm was associated with the *B*. *cereus* isolate due water and/or protein content (Dubois *et al.*, 2005). This observation was the same for the bands at 1313, 1345 (lipids) and 1890 nm (carbohydrate). The results are in agreement with Kammies, *et al*., (2016), that reported a similar phenomenon where the Gram-negative isolates were located in the positive region of the loading line plots and the Gram-positive isolates in the negative region (Kammies *et al.*, 2016). A summary of the absorption band assignment associated with the individual bacteria can be seen in Table 4.1.

Table 4. 1 Wavelengths of interest and absorption band assignment for respective loadings of **Group 1** and **2**

PC	Wavelength (nm)	Absorption band assignment	Cell wall component	Associated bacteria
		Group 1		
PC1	1907	O-H & NH_2 stretch, 1st overtone & combination band region	Water/ protein	*E. coli*
	1868	O-H & C=O stretch	Carbohydrate	
PC 2	1160, 1127, 1291	C-H stretch, 2nd overtone		*S. aureus* 1, *B. cereus*
	1307, 1378	C-H stretch & deformation	Lipid	*S. aureus* 1
	1416	N-H stretch, 1st overtone (CONH)	Protein	
	1885	O-H & C=O stretch	Carbohydrate	*S. aureus*
PC3	980	O-H stretch, 2nd overtone	Water	*S. aureus* 1
	1116, 1149	C-H stretch, 2nd overtone	Carbohydrate	*S. aureus*
	1307, 1345, 1378	C-H stretch & deformation	Lipid	*E. coli*
	1852, 1855	O-H & C=O stretch	Carbohydrate	
		Group 2		
PC 1	980	O-H stretch, 2nd overtone	Water	*B. cereus*, *Salmonella*
	1122, 1160	C-H stretch, 2nd overtone	Carbohydrate	*E. coli*
	1340, 1372	C-H stretch & deformation	Lipid	*B. cereus*, *Salmonella*
	1863	O-H & C=O stretch	Carbohydrate	*B. cereus*, S. *Salmonella*

	1907	O-H bend 2nd overtone and deformation combination, N- H stretch, 2nd overtone	Water/ protein	*E. coli*
PC 2	1122, 1160	C-H stretch, 2nd overtone		*B. cereus*, *S. aureus*
	1378	C-H stretch & deformation	Lipid	
	1416	N-H stretch, 1st overtone (CONH)	Protein	
	1885	O-H & C=O stretch	Carbohydrate	*S. aureus*
	1907, 1923	O-H & NH_2 stretch, 1st overtone & combination band region	Water/ protein	*E. coli*, *B. cereus*
PC3	980	O-H stretch, 2nd overtone	Water	*E. coli*
	1122, 1154	C-H stretch, 2nd overtone	Carbohydrate	*S. aureus* 1
	1313, 1345, 1372	C-H stretch & deformation	Lipid	*B. cereus*, *S. aureus* 1, *S. epidermidis*
	1411	N-H stretch, 1st overtone (CONH)	Carbohydrate/Protein	*S. aureus* 1
	1852, 1890	O-H & C=O stretch		*Salmonella*
	1923	O-H & NH_2 stretch, 1st overtone & combination band region	Water/ protein	*B. cereus*

Group 3 PCA results of bacteria – NA & TSA

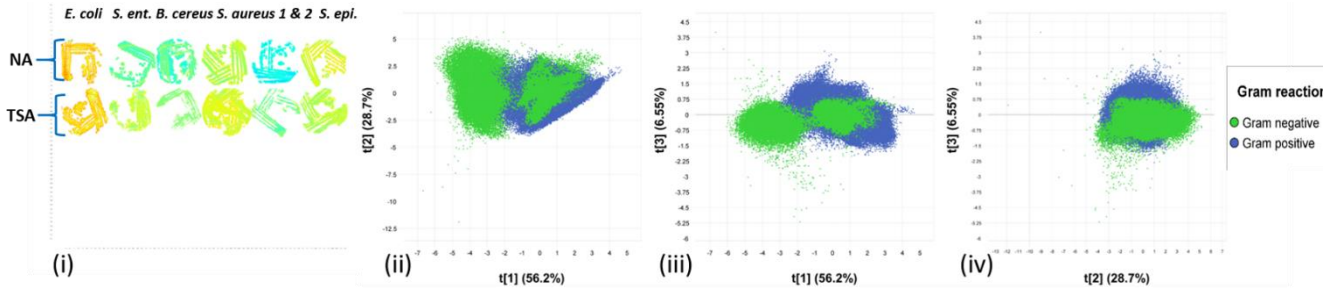

Figure 4.10 Pixel-wise PCA of Group 3 illustrating poor separation between the Gram-positive and Gram-negative isolates. The images depict the scores in descending order of principal components (PC's). PCA score image (i), shaded according to the density of the samples (LTR: *E. coli*, *S. enteritidis*, *B. cereus*, *S. aureus*1, *S. aureus*2 and *S. epidermidis*). PCA score plot (ii) of PC1 (56.2%) against PC2 (28.7%) shaded according to the two classes Gram-positive bacteria and Gram-negative bacteria. PCA score plot (iii), of PC1 (56.2%) vs. PC3 (6.55%) shows some separation for the Gram-negative bacteria. In PCA score plot (iv), PC2 (28. 7%) vs. PC3 (6.55%), shows even less separation.

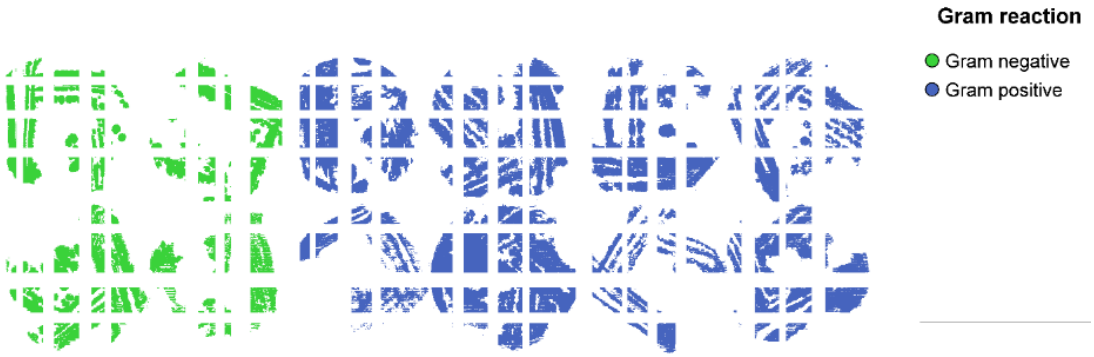

Figure 4.11 Objects created by segmentation of bacterial samples into approximately 25 objects per sample, shaded as per the legend (Total number of objects 170).

In **Group 3** bacteria are now classed into their respective Gram stain reactions (Gram-positive or Gram-negative). This was done in order to evaluate the effect of the two different media on the separation of the bacterial isolates based on the two major classes. (Results for the individual bacteria can be seen in **Addendum A**, Figure A1).

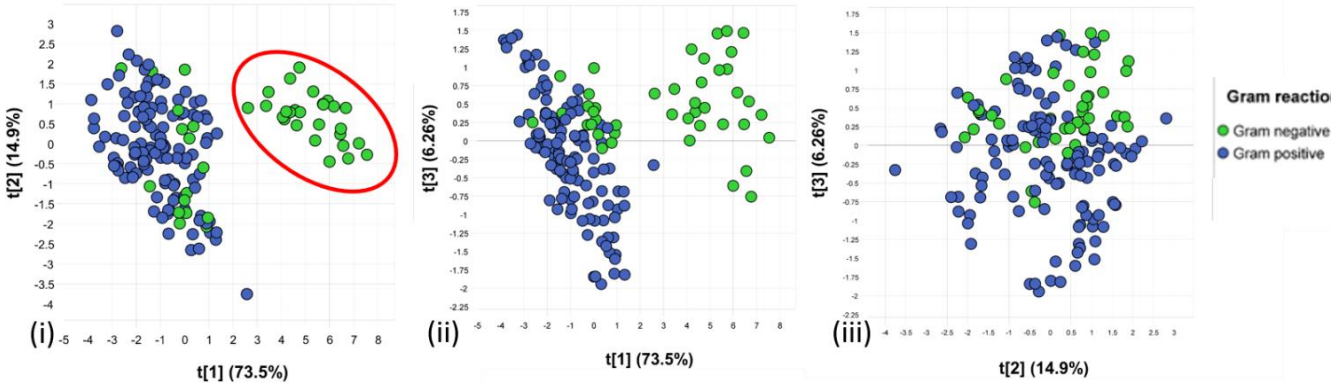

Figure 4.12 Object-wise mosaic illustrating improved separation between the bacterial isolates. The isolates were grouped together as per the legend. The PCA score plot (i), of PC1 (73.5%) vs. PC2 (14.9%) *E. coli* isolates clustered together, showing separation is possible. In PCA score plot (ii), in the direction of PC1 (73.5%) vs. PC3 (6.26%), a similar observation was made. In PCA score plot (iii), PC2 (14.9%) vs. PC3 (26.26%), the samples were scattered in a random manner – no clear separation observed.

When comparing the mosaics of the pixel-wise PCA score plots to that of the object-wise PCA score plots, an improved separation between the bacterial isolates was observed. The variance in PC1 (73.5%) vs. PC2 (14.9%) (Figure 4.12(i)) increased by 17.3%. In PC2 vs. PC3 a decrease of less 13.8% in the variance is observed. This shows that PC1 accounts for a major fraction of the total variance in the sample. Fewer data points are used in the PCA computation which is the reason for the improved separation in the samples. The score plot shown in Figure 4.12(i) & (ii) clearly separates the gram-negative *E. coli* samples from the rest of the isolates, which when compared to the score plots generated by the pixel-wise mosaics was a large improvement as no clear separation is observed in Figure 4.10. The *Salmonella* sample however was poorly separated and overlapped entirely with the Gram-positive isolates. Overall the pixel-wise mosaic showed very poor separation, therefore it was decided to continue with the object-wise approach for the remainder of the analysis.

When closely examining the score image of both pixel- and object-wise approaches, the differences observed in the colour densities of the bacteria was confirmed. In Figure 4.10(i), the *E. coli* samples were clearly very different from the rest (orange). Cool colours (blue, green and cyan) relates to low-density scores, whereas warmer colours (red, orange and yellow) relate to higher density scores (Feng and Sun, 2014). This confirms that the chemical composition of the *E. coli* samples is considerably different from the rest of the isolates.

Although the overall outer membranes of bacteria are similar, the Gram-positive bacteria contain a thick peptidoglycan layer, whereas the Gram-negative bacteria have a very thin peptidoglycan layer which is one of the main differences observed in bacteria. When comparing *E. coli* and *Salmonella*, Salton & Kim (1996), reported that the inner and outer membranes of *E. coli* bind to each other at several sites (Beyer patches) and therefore causes a break in the peptidoglycan layer. This could be the reason for the differences observed when comparing these two pathogens.

Loadings Group 3

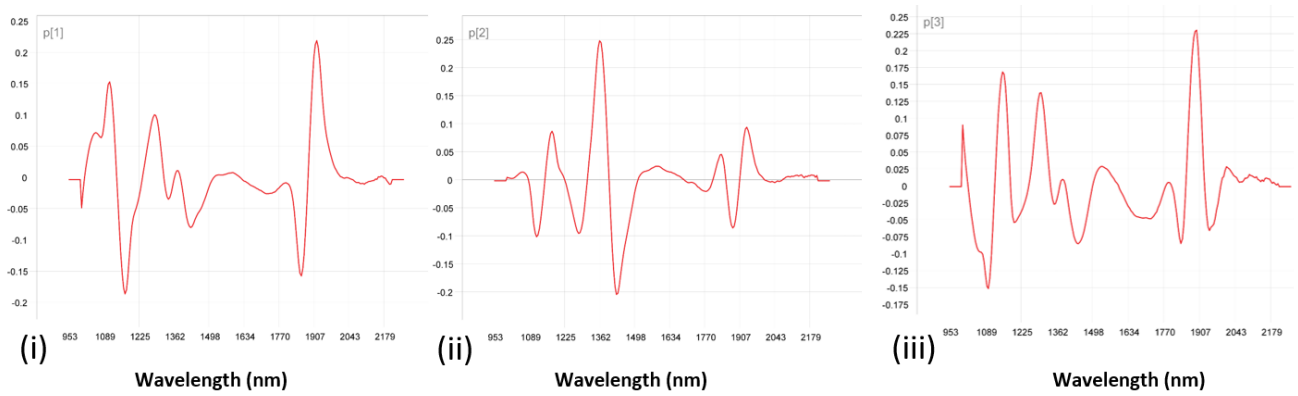

Figure 4.13 PCA loading lines of **Group 3** (all six isolates). Important bands were observed for (i) PC1 at 1111, 1285 and 1918 nm (pos. region) and 1171, 1340, 1427 and 1858 nm (neg. region). For (ii) PC2 at 1176, 1362, 1836 and 1934 nm (pos. region) and in the negative region at 1116, 1280, 1427 (inversed in PC1) and 1879 nm. In (iii) PC3 (pos. region) prominent peaks are observed at 1007, 1154, 1302, 1531 and 1896 nm. In the positive region the prominent peaks are at 1089, 1498, 1836 and 1945 nm.

The loading lines for **Group 3** can be seen in Figure 4.13. The prominent peak in the positive region for PC1 observed at 1918 nm is related to the O-H and NH_2 stretch in the first overtone and combination band region and was attributed to moisture (Osborne, 2000). When re-examining the PCA score plots in conjunction with the loading line plot of PC1 vs PC2 (Figure 4.12(i)), the *E. coli* samples that clustered together separating it from the other isolates in **Group 1** is based on moisture absorption at this band. This could be as a result of different water binding capacities bacteria exhibit in the NIR region. It is also possible that this is due to protein content in the cell walls of Gram-negative bacteria (Navarre and Schneewind, 1999). However, water condensate accumulates on the inside of the petri dish due to the respiring organisms and this could be another reason for the water-spectra observed (Brewer & Carski, 1950). Respiration of bacteria occurs when in a nutrient-rich environment (growth media) and therefore results in the accumulation of water in the petri dish. Another contribution to the formation of water vapour could be that the bacteria is sensitive to temperature fluctuations. The samples experienced a decrease in temperature from 37 °C after incubation to ambient temperature at 23 °C prior to imaging. The remaining bands attributed to water in the sample was observed in both PC1 and PC2 at 1427 nm

which relates to the C-H stretch in the first overtone (Osborne, 2000), this is similar for the band at 1443 nm in PC3. Again in PC2, the band at 1934 nm and in PC3 at 1945 nm is associated with the O-H bend and N-H vibrations in the second overtone and combination band region (Wilson *et al.*, 2015; Bezuidenhout, 2018).

The remaining wavelengths of interest observed in PC1 were at 1111 and 1171 nm which is as a result of the C-H stretching in the second overtone (Genot *et al.*, 2014) which could be due to carbohydrate content. At 1285 nm and 1340 nm, two peaks related to the C-H stretch in the second overtone is observed which was ascribed to lipid content (Osborne, 2000; Mosier-Boss, 2017). Lastly, a sharp peak at 1858 nm is due to the O-H stretch and C=O stretch and was attributed to carbohydrate content (Osborne, 2000). The Gram-positive bacteria is strongly associated with the negative region in the PCA score plots. The *E. coli* isolate was represented in the positive region in the score plots.

In PC2 (Figure 4.13(ii)), the Gram-positive isolates were represented in both positive and negative region in the PCA score plots. The bands observed at 1176 and 1116 nm was due to carbohydrate content in the sample. We observed two bands at 1280 and 1362 nm and attributed it to lipid content in the Gram-positive cell. A peak was observed at 1836 nm (pos. region) indicating the presence of a carbohydrate in the Gram-positive cell. In PC 3 (Figure 4.13(iii)) a sharp peak is observed at 1007 nm is due to the O-H and N-H vibrations in the second overtone due to the carbohydrate and protein absorption respectively. The absorption bands around 1100 – 1200 nm is related to $C\text{-}H_2$ and $C\text{-}H_3$ vibrations in the second overtone due to lipids. The peaks at 1302, 1389 and 1351 nm are due to the C-H stretch and deformation also due to lipids (Kammies, 2018). The band at 1498 nm is characteristic of a protein due to the intramolecular bond in the N-H functional group (Tachuck, 1987; Osborne, 2000; Rodriguez-Saona *et al.*, 2001). Lastly, the peak observed at 1836 nm was attributed to carbohydrate, whereas the band at 1945 nm was attributed to protein or water content in the sample. The *E. coli* samples were clearly separated from the rest of the isolates and strongly associated with the positive region in PCA score plot of PC1 (73.5%) vs. PC3 (6.26%) (Figure 4.13(ii)). The *Salmonella* isolates, however, overlaps almost entirely with the Gram-positive bacteria and showed poor separation.

Overall comparison

When comparing the object-wise results for **Group 1**, **2** and **3** (Figures 4.4, 4.7, 4.12), the PCA models produced quite similar results. In **Group 1** 76.3 % of the total variance in the sample is accounted for by PC1. In **Group 2** and **3,** 74.4% and 73.5% is accounted for by PC1. It would appear that the plates containing the bacterial isolates on nutrient agar, provided better separation of the individual bacteria overall. This can be seen in the PCA score plots of Figure 4.4. The *S. aureus* and *E. coli* isolates are slightly separated from the rest of the isolates. The samples (data points) are also packed more closely together in **Group 1** compared to the other groups where it

is more spaced out. This observation, however, has to be confirmed as the model containing both NA and TSA plates showed separation of the Gram-negative isolate, *E. coli* on both NA and TSA.

Group 4: Pathogenicity

Pathogenic and non-pathogenic bacteria differ in the fact that pathogenic bacteria require structures such as protein A to adhere to host cells (Nakakimura *et al.*, 2012). This unique characteristic is therefore worth investigating, especially in organisms within the same genus. To evaluate the pathogenicity of foodborne pathogenic bacteria, we selected two pathogenic *Staphylococcus aureus* species (ATCC 29213 & ATCC 25923) to compare to the non-pathogenic *Staphylococcus epidermidis*. The organisms were streaked out onto both NA and TSA plates to observe the overall effect of the growth media on the NIR spectra.

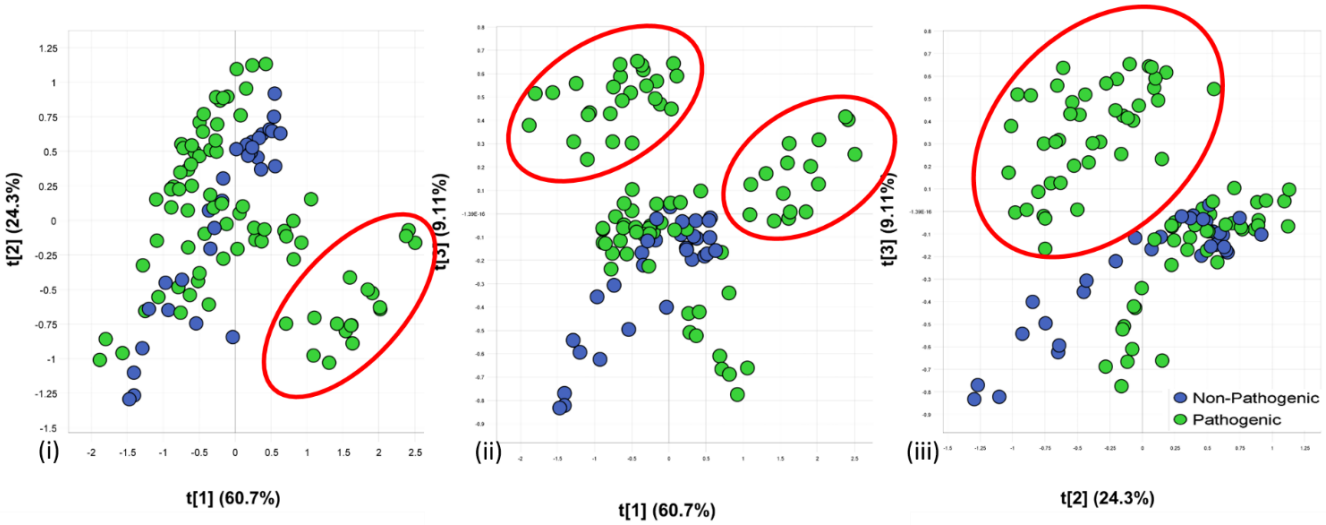

Figure 4.14 Object-wise mosaic illustrating separation between the pathogenic vs. non-pathogenic isolates. The PCA score plot (i), of PC1 (60.7%) vs. PC2 (24.3%) some clustering observed for the pathogenic bacteria, showing distinction is possible. In PCA score plot (ii), in the direction of PC1 (60.7%) vs. PC3 (9.11%), a similar observation was made for both pathogenic isolates. In PCA score plot (iii), PC2 (24.3%) vs. PC3 (9.11%), pathogenic bacteria clustering together.

In the direction of PC1, 60.7% of the total variance in the sample was accounted for. The *S. aureus*1 isolate on both NA and TSA plates clustered together as highlighted in PCA score plot of PC1 vs. PC2 (Figure 4.14(i)). In PC1 (60.7%) vs. PC3 (9.11%), Figure 4.14(ii), both pathogenic *Staphylococcus* species streaked out on NA formed two distinct clusters, showing that the media may indeed have an effect on the separation (or lack thereof) of bacteria.

This observation was further amplified in PC2 (24.3%) vs.PC3 (9.11%). In the direction of PC3 (9.11%), the pathogenic species that were cultured on NA media clustered together. Overall the non-pathogenic isolates were

poorly separated using an object-wise analysis. Examining the loading lines (Figure 4.15) for **Group 4**, important wavelengths contributing to the variability were observed in both positive and negative regions.

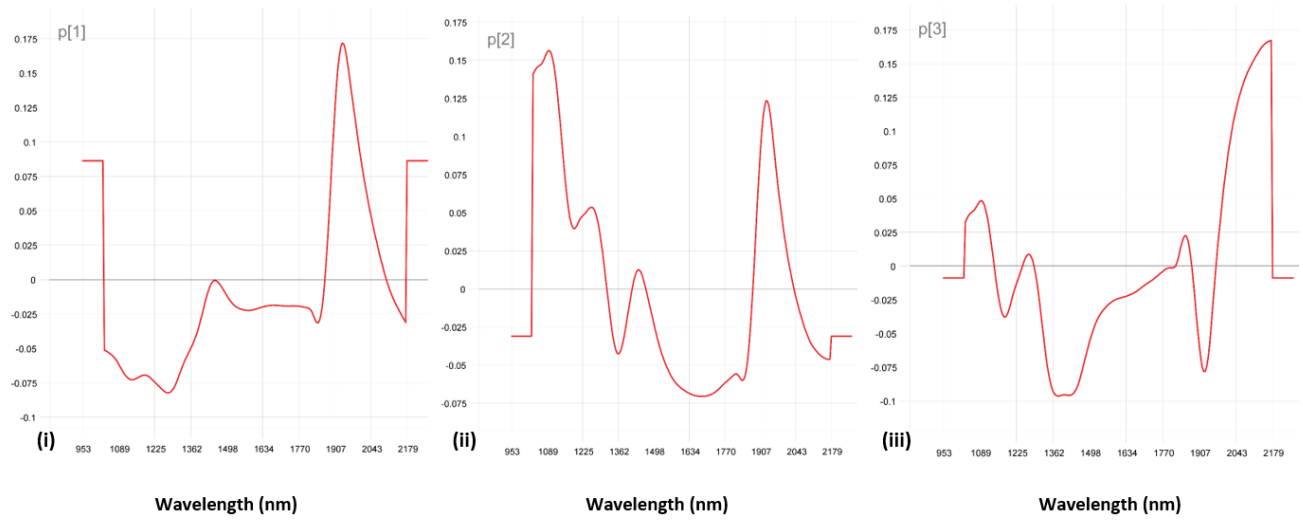

Figure 4.15 PCA loading lines of Group 2 (pathogenic vs. non-pathogenic isolates). Important absorption bands were observed for (i) PC1 at 1934 nm, (ii) PC2 at 1383 and 1907 nm and (iii) PC3 with bands at 1433, 1907 nm in the positive region (inversed in PC2) and at 1672 nm.

In Figure 4.15(i) loading plot 1 shows a prominent absorption band at 1934 nm due to moisture in the sample. Dubois *et al*., (2005), found that the band at 1934 nm (O-H stretch combination and deformation) was useful to distinguish bacteria on food-specific cards. The research group ascribed the separation to the effect of moisture and hydrogen bonding in the sample matrix. When examining the PCA score plot of PC1 (60.7%) vs. PC2 (24.3%) (Figure 4.14(i)), majority of the variance is accounted for in the direction of PC1, which is where a cluster formed by the pathogenic isolates on NA is located. In loading line (ii) PC2, two prominent peaks were observed in the negative region of the loadings at 1383 and 1907 nm. When examining the PCA score plots in the direction of PC2 (24.3%), the pathogenic isolate (*S. aureus*1) as well as the non-pathogenic isolate *S. epidermidis*, is located in this region. This confirms the inter-genera similarities of the bacteria. The band at 1383 nm is related to the C-H stretch in the second overtone and could be as a result of carbohydrate content (Osborne, 2000; Mosier-Boss, 2017). The band at 1907 nm in the negative region in PC2 was also observed in the positive region in PC3 and is as a result of water absorption. This band represents the water absorption properties of both pathogenic isolates. Similarly, the band at 1433 nm ((ii) PC3) is as a result of moisture content. Lastly, the band at 1672 nm (CH_3) in the negative region is a contribution from the lipid content contained in the non-pathogenic cell wall (Osborne, 2000; Dubois *et al.*, 2005; Kammies, 2018).

Different culture media contain different nutrients required to support bacterial growth (Mlynáriková *et al.*, 2015). Biomolecules such as carbohydrates, proteins, lipids and nucleic acids are contained in the culture

media. Nutrient agar is a non-selective agar medium that contains peptone, beef-or yeast extract and bacteriological agar powder (Neal, 2019). Whereas tryptic soy agar is another form of nutrient agar and contains casein and soybean meal. Both agar mediums are suitable for culturing a vast number of microorganisms. Each bacterial cell will exhibit a different intake of the nutrients provided which will cause the bacteria to have different cell and colony compositions (Mlynáriková *et al.*, 2015). When examining the overall PCA scores, it suggests that the bacteria grown on NA clustered together which is less prevalent in the TSA samples. It would seem that NA could be the best suitable medium for pathogenic bacteria (Witkowska *et al.*, 2019). However, this observation is unclear whether it is the growth media that is the reason for the separation or if it is the actual chemical composition of the pathogenic bacteria that is responsible for the variability. These differences could potentially affect the spectra of the different pathogenic bacteria.

When examining the average spectra for the bacterial isolates (**Addendum A,** Figure 4A), we found that the overall shapes were quite similar, with the TSA samples generally having increased peak definition in certain areas. The overall results suggest that the growth medium does indeed have an effect on the spectra, however, does not hinder the separation too greatly. This observation coincides with a study conducted by Leroux *et al.*, (2014). The researchers compared the effect of three different growth media (Columbia agar, Tryptic soy agar and Mueller Hinton 2 agar) for the differentiation of *Escherichia coli*, *Staphylococcus aureus*, *Pseudomonas aeruginosa*, and *Enterococcus faecalis*. The researchers found that although the agar had a strong contribution to the overall shape of the spectra, it had no effect on the differentiation of the organisms.

4.1.3 Partial Least Squares Discriminant Analysis (PLS-DA)

PLS-DA models were calculated for the object-wise mosaics of **Group 1** - **4**. The overall performance of the models was determined by calculating the classification accuracy, sensitivity, specificity and misclassification rate. A summary of the overall performance of the models can be seen in Table 2.

Table 4. 2 The performance measures used to assess the overall object-wise PLS-DA results for the different foodborne pathogenic bacteria.

Group	Optimal number of LV's	Classification Accuracy (%)	Sensitivity (%)	Specificity (%)	Misclassification Rate (%)
1	8	99.55	97.33	100	0.45
2	9	100	100	100	0
3	5	98.29	100	96.64	1.71
4	7	98.72	100	97.47	1.28

LV = Latent variables

Group 1 and **2** obtained a classification accuracy of 99.55 and 100% respectively. Proving once more, that although using different growth media may produce a shift in intensity in the individual spectral results, it does not hinder the separation of foodborne bacteria too greatly. In **Group 3** a decrease in the classification accuracy can be seen. The model obtained a classification accuracy of 98.29%, which when compared to **Group 1** and **2** is not too large of a decrease. One reason for this could be that in the combined model, more variation is contained compared to the mosaics that contained only NA or TSA samples. For the pathogenicity model, a classification accuracy of 98.72% was obtained. However, this model cannot directly be compared to the rest of the models as the objective was quite different from that of the previous models. Here, we selected only *Staphylococci* species to distinguish pathogenic species from non-pathogenic species within the same genera.

The sensitivity is simply the rate of the true positives (TP) meaning the number of correctly predicted samples. **Group 1** obtained a sensitivity of 97.33%. The sensitivity for **Group 2**, **3** and **4** was 100%. This indicates that the individual models have a high probability of correctly predicting the individual bacterial isolates. The specificity refers to the probability of the remaining samples to be correctly predicted as true negatives (TN). All the models obtained a specificity percentage above 96.46% (**Group 3**). For **Group 1** and **2,** a specificity of 100% was obtained, whereas **Group 4** obtained a specificity of 97.47%. The lowered specificity rate obtained for **Group 3** can again be ascribed to the large variation contained within the data set. The misclassification rate of all the

models was quite low with **Group 2** obtaining the lowest percentage misclassification (0%) and therefore being the model with the best overall performance. Overall, the statistical performance of the models was excellent, showing that irrespective of the growth media used, the models could be used to distinguish between the individual bacterial isolates. A visual representation of the overall performance of the models can be seen in the prediction maps below, Figure(s) 4.16 & 4.17.

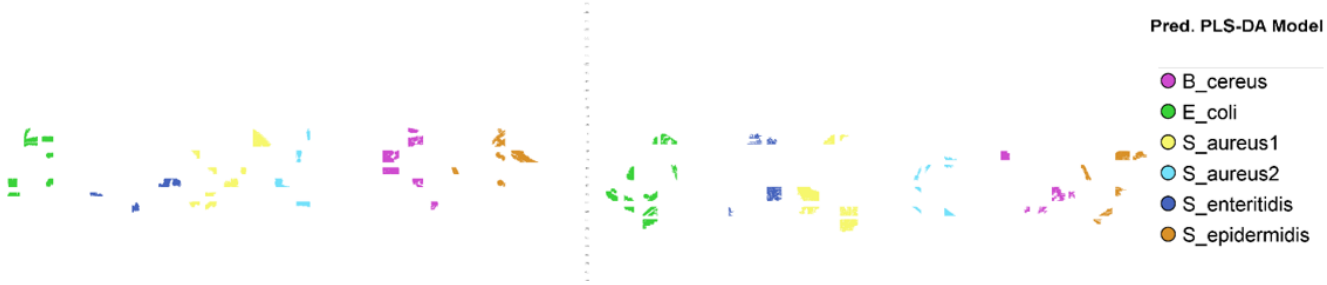

Figure 4.16 Prediction maps for (i) **Group 1** and (ii) **2** showing object-wise analysis applied to bacterial isolates contained in the test set. The bacterial isolates were classed as per the individual isolates.

It can be noted that all the objects contained in the test set were correctly classified. The calibration model contained enough variation (70%) to account for all the samples contained in the test set (30%). The isolates on both NA and TSA were correctly classified, confirming the observations made in the object-wise PCA score plots of the same models (Figure 4.4 & 4.7). Using different growth media resulted in a slight amplitude change in the spectral results (**Addendum A**, Figure A2 & A3). However, had no effect on the separation of the organisms.

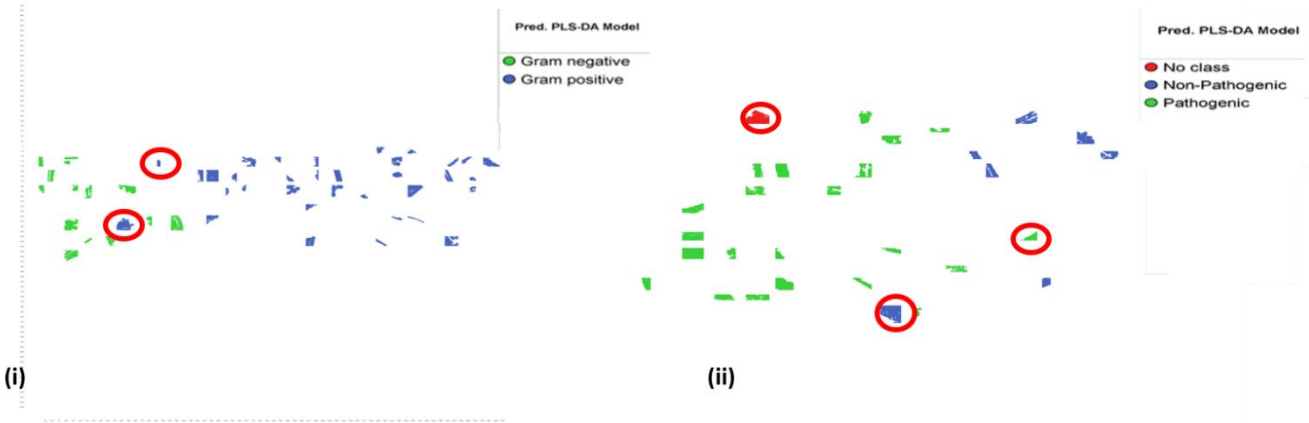

Figure 4.17 Prediction maps for **Group** (i)**3** and (ii)**4** showing object-wise analysis applied to bacterial isolates contained in the test set. The highlighted objects in **Group 3** were misclassified as Gram-positive (blue) where in actual fact they were Gram-negative (green) samples. In **Group 4** two objects were incorrectly classified as non-pathogenic (blue). One object was misclassified as pathogenic and one object was not classed at all and assigned to 'no class' (red).

When examining the digital images of **Group 3 (ADDENDUM A**, Figure A6), the objects that were misclassified belonged to the *Salmonella* isolate on both growth media. We noted that the *Salmonella* isolate streaked out on NA contained a thicker streak on the agar surface, compared to the rest of the streaked-out organisms (**Addendum A**, Figure A6 & A7). The growth areas were clumped together, rather than being separated by each streak. This suggests that the thickness of the colonies plays a role in separating each individual organism. It also suggests that the incubation time resulted in the overgrowth of the organism which is the reason for the clump formation. This observation correlates with the study done by Le Roux *et al*., (2015). The researchers used three different growth media to examine bacterial growth with varied incubation times and concluded that different incubation periods will result in a varied thickness within colonies.

Examining the digital image of the same organism streaked out on TSA (**Addendum A**, Figure A6), the inverse is observed. The plates contained only a thin layer of colonies that were formed non-uniformly. One reason for the misclassification could be that the colonies were mixed with some of the agar, which resulted in mixed spectra of the organism along with the agar. Maquelin (2000) and co-workers observed a similar phenomenon for the identification of bacteria using Raman spectroscopy on solid culture media (Maquelin *et al.*, 2000). The lack of thickness in the colonies contributes greatly to the NIR signal response and results in mixed spectral results. The researchers used a vector correction method to subtract the combined signal (agar and bacteria), that could not be distinguished from the pure signal of the culture media. A PCA model was then used, followed by supervised linear discriminant analysis (LDA), which obtained 100% classification accuracy for the training set and 83% for the test set.

In the object-wise analysis, the average spectra are obtained for each pixel to be used in the prediction procedure. It could be that in the average spectra for that specific object, the ratio of the pixels representing the agar was more than that of the bacterial colony.

PLS-DA VIP Scores

The main wavelengths contributing to the overall performance of the PLS-DA models were found in the VIP scores (Figure 4.18 & 4.19).

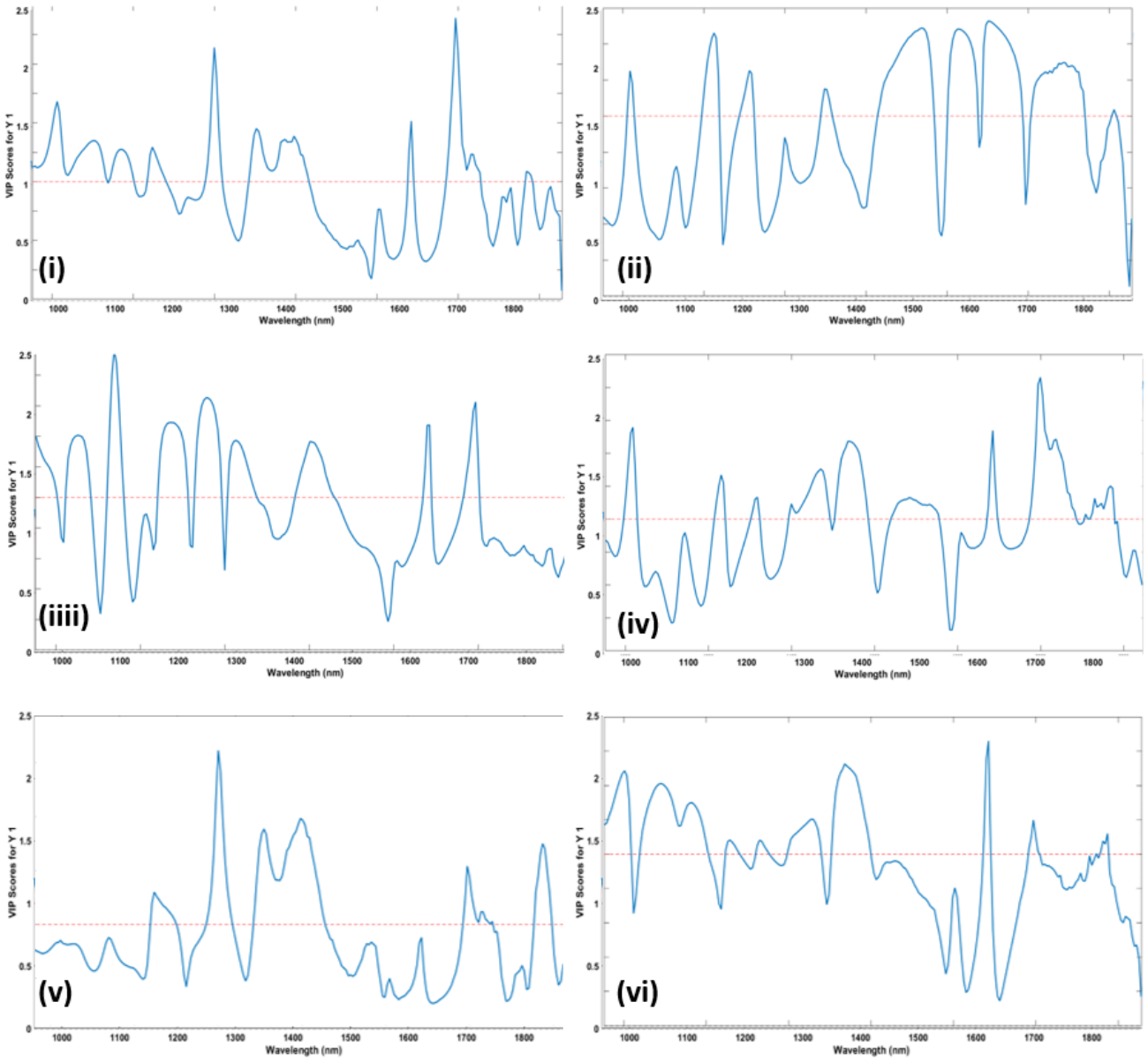

Figure 4.18 VIP scores (8 LV's), for object-wise PLS-DA model for **Group 1**. The prominent bands located above the threshold line (red) as indicated were at (i) 1013, 1105, 1171, 1247, 1400, 1503, 1596, 1885, 1994, 2038,2174 and a valleyed trough at 1574 nm was observed; (ii) 1018, 1225, 1313, 1503, 1738, 1819, 1907, 2076 and 2212 nm; (iii) 1056, 1138, 1269, 1356, 1427, 1601, 1885 and 1994 nm; (iv) 1081, 1231, 1318, 1400, 1465, 1536, 1689, 1885, 1999, 2038, 2130, 2174 nm; (v) 1247, 1400, 1509, 1596, 1994, 2031 and 2174 nm; (vi) 1002, 1089, 1160, 1258, 1329, 1460, 1536, 1885, 1994, 2163 and 2174 nm.

The variable importance in projection (VIP) scores, is a measurement of the most important wavebands contributing to the variation observed in the predictor variable (X) as well as the response variable (Y) (Xu *et al.*, 2018). Since the measurement is similar to that of the loading lines, similar observations were made with regards to important wavelengths contained in the sample. The weighted sum of squares of the PLS weight is used in the calculation, taking into account the amount of explained Y variance which is contained in each extracted latent variable (dimension) (Farrés *et al.*, 2015). Calculating the average of the squared VIP scores, produces a total sum of squares equal to one. This is used as a threshold to determine which variables are greater than one, which then indicates the most significant wavebands (Xu *et al.*, 2018)

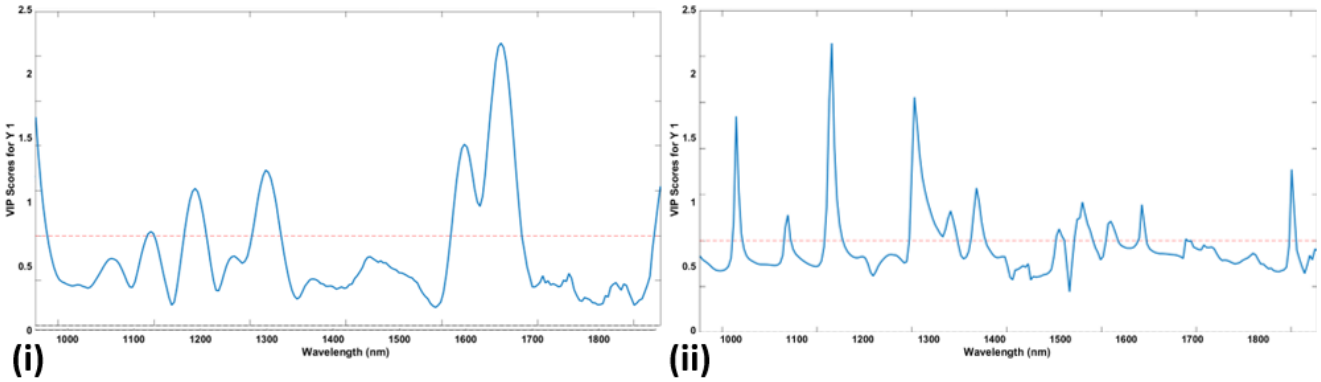

Figure 4.19 VIP scores for **Group 3** (LV = 5) and **4** (LV = 7) for VIP score Y1 indicating prominent absorption bands. Important wavelengths for these models were observed at (i) 953, 1193, 1285, 1433, 1847, 1923, 2256 and a trough at 1907 nm was observed; (ii) 1029, 1138, 1231, 1405, 1482, 1536, 1711, 1760, 1820, 1885 and 2201 nm.

VIP scores for **Group 3** (Gram-stain reaction) and **4** (pathogenicity) were identical for VIP score Y1 and Y2 and are therefore only reported once for each group. The distinguishing factor for the Gram-stain reaction model was whether a particular bacterium is Gram-negative (Y1) or not (Y2). Similarly, for the pathogenicity model Y1 represents non-pathogenic bacteria and Y2, pathogenic bacteria. Majority of the wavebands observed here have been reported previously (Table 4.1), in addition to this, Table 4.3 depicts the remaining important wavebands found in the PLS-DA models.

Table 4. 3 Wavelengths of interest and absorption band assignment for respective loadings of **Group 1**, **3** and **4.**

VIP score	Wavelength (nm)	Absorption band assignment	Chemical component
		Group 1	
1	2038	O-H stretch combination	Carbohydrate
	2174	O-H & N-H combination	Protein
2	2076	N-H sym. Stretch & amide II	Protein
	2212	O-H & N-H combination	Protein
4	2130	N-H stretch, C=O	Protein
5	2032	O-H stretch combination band region	Carbohydrate
6	2163	2 (amide I & III)	Protein
		Group 3	
1	2256	CH_2 bend & stretch combination	Carbohydrate
		Group 4	
1	2201	O-H & N-H combination	Protein

4.2 Evaluating the effect of different plating techniques on the NIR spectra of bacteria

4.2.1 Spectral Analysis

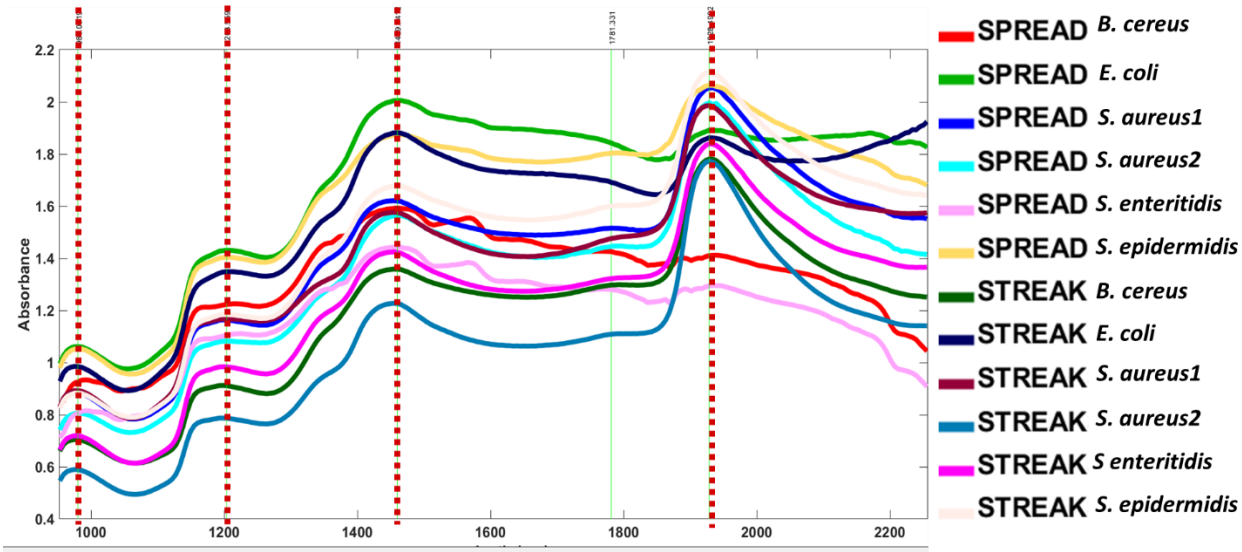

Figure 4.20 Average unprocessed and spectra obtained for individual bacteria on NA using streak and spread plate technique.

As discussed previously, the overall spectral results were quite similar. However, the spread plates of the *B. cereus* and *S. enteritidis* isolates deviated from the spectral shape obtained from the rest of the isolates. The deviation occurs at 1928 nm, which is associated with moisture contained in the sample. When examining the digital images of these two plates, it was evident that a large area of overgrowth had occurred where the initial inoculum was made (**ADDENDUM B**, Figure B21).

Data Analysis

4.2.2 Principal component analysis (PCA)

Model optimization

For the first objective the pre-treatments applied to all the models were kept constant as it provided the best separation between the organisms. However, it is important to note that a 'one fits all' approach cannot always be implemented. Different pre-treatments were therefore implemented and compared to determine which would provide more defined clustering and yield optimal results. PCA was calculated before applying various pre-treatments: **a**) SNV, **b**) SNV and Savitzky-Golay (2nd derivative, 3rd polynomial, 9 pt. smoothing filter), and **c**) Savitzky-Golay (2nd derivative, 3rd polynomial, 5 pt. smoothing filter) to mean centred data, separately. To

illustrate this, we will focus on the Gram-positive bacteria using a streak plate technique, the Gram-negative bacteria using a spread plate technique and lastly a combination of the bacteria using both techniques.

A combination of SNV and SG proved to be the best pre-processing technique for **Group 5** (Figure 4.21). When examining the score image of **Group 5**, (i), before pre-processing was applied, it suggests that the *B*. *cereus* and *S*. *aureus* 2 isolates are slightly different from the remaining two isolates. After pre-processing was applied, the inter-class (Gram-positive) similarities within the organisms are highlighted. All the images now clearly exemplify high score values (red, orange or yellow), showing that the most important biochemical components within the Gram-positive cell are similar.

Group 5: PCA results of Gram-positive bacteria – *streak*

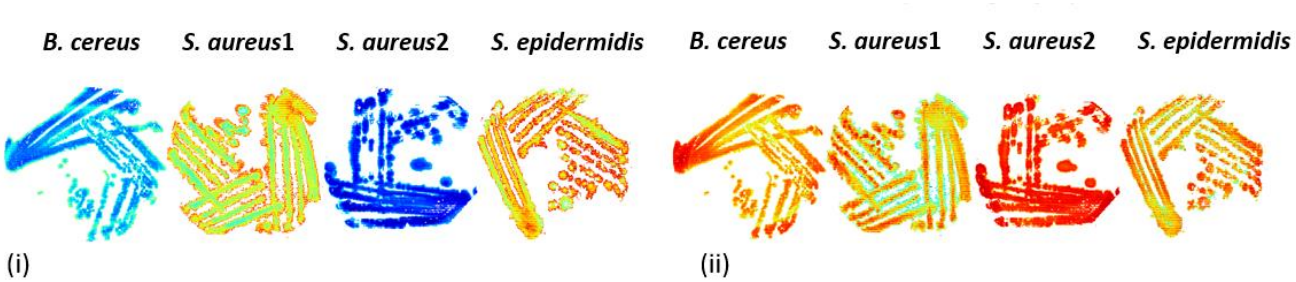

Figure 4.21 Illustration of score images (i) before and (ii) after pre-processing (SNV & SGd2; (9pt.)) is applied.

The separation between the individual bacteria of **Group 5** is observed in Figure 4.22 In PCA score plot (i) of PC1 (45.4%) vs. PC2 (42%), all the isolates form four distinct clusters with densely packed data points. In PCA score plot (ii) of PC1 (45.4%) vs. PC3 (5.94%) clustering is observed for *S*. *aureus*2, with densely packed data points. The data points of the remaining isolates are more spaced out, with some overlapping between *B*. *cereus*, *S*. *aureus*1 and *S*. *epidermidis*. Finally, in PCA score plot (iii) of PC2 (42%) vs. PC3 (5.94%), the *B*. *cereus* isolate form a more spaced out cluster, whereas the remaining isolates overlap.

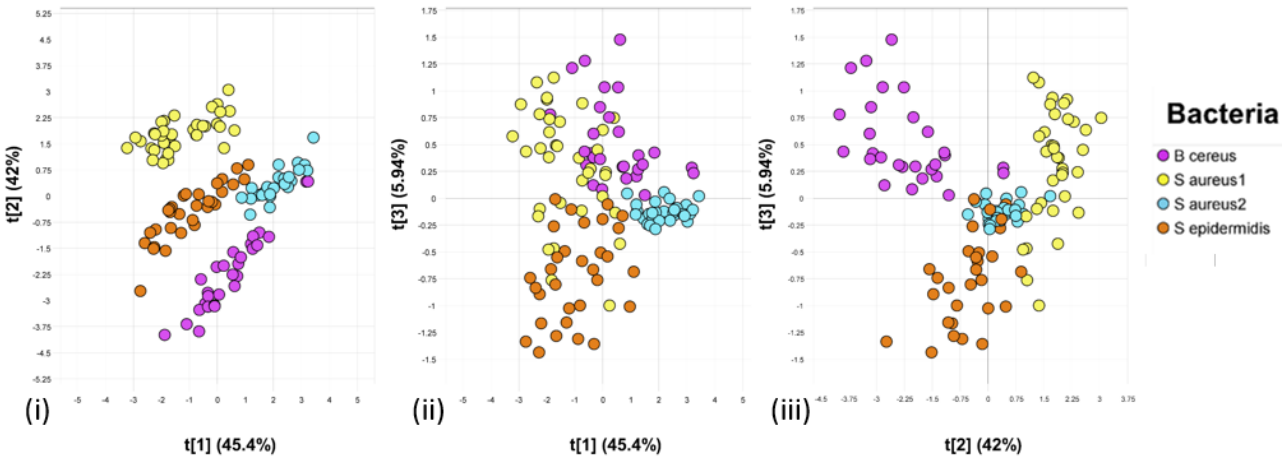

Figure 4.22 PCA score plots illustrating separation of the Gram-positive isolates treated with a combination of SNV and SGd_2; (9pt.). The PCA score plot (i), of PC1 (45.4%) vs. PC2 (42%). In PCA score plot (ii), of PC1 (45.4%) vs. PC3 (5.94%), some clustering observed for *S*. *aureus*2, with some overlapping. In PCA score plot (iii), PC2 (42%) vs. PC3 (5.94%), some clustering observed for *B*. *cereus*, *S*. *aureus*1 and *S*. *epidermidis*.

Loadings Group 5

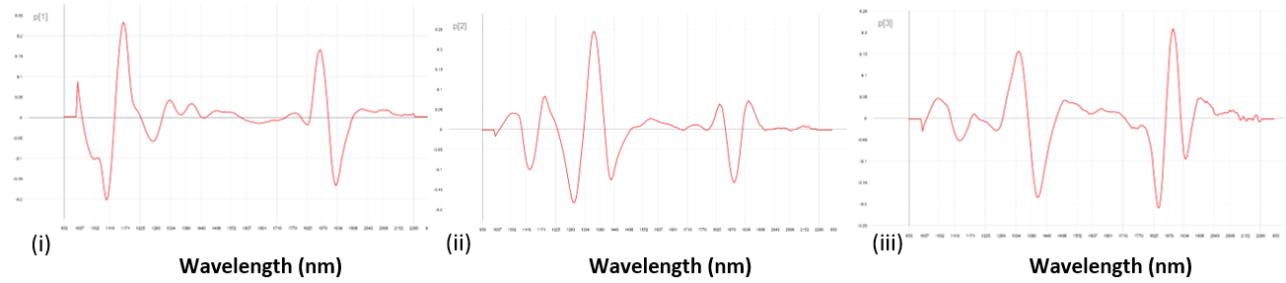

Figure 4.23 PCA loading lines of **Group 5**. Important wavebands were observed for (i) PC1 980, 1165, 1874 nm (pos. region) and at 1105 and 1928 nm (neg. region). For (ii) PC2 at 1367 nm (pos. region) and in the negative region at 1291 and 1432 nm. In (iii) PC3 two prominent bands are observed in the positive region at 1345 and 1896 nm and in the negative region at 1416, 1847 and 1939 nm.

When evaluating the PCA score plots and loading lines for **Group 5** three prominent peaks are observed in the positive region of loading line 1 (Figure 4.23(i)). The band at 980 nm is responsible for the clustering observed for *B. cereus* in the PCA score plot of PC1 vs. PC2 (Figure 3(i)). The band is related to the O-H stretch in the second overtone, attributed to the absorption of water (Osborne, 2000; Clevers and Kooistra, 2006). A second positively loaded peak is observed at 1165 nm, which is related to the C-H stretch in the second overtone accounting for the lipid content (Osborne, 2000; Genot *et al.*, 2014). Lastly, the band at 1874 nm is associated with the carbohydrate content found in the cell wall of *B. cereus*. The band is related to O-H and C=O stretch in the first overtone and combination band region (Osborne, 2000). The bands observed at 1105 nm, shows the separation

of *B. cereus* from the rest of the isolates which is based on the two negatively loaded bands at 1105 nm (carbohydrate) and 1928 nm (moisture). Some of the *S. epidermidis* samples (data points) are also observed in this region, which shows inter-class (Gram-positive) similarities.

In Loading 2 (Figure 4.23(ii)), a positively loaded peak at 1367 nm is responsible for the presence of the *S. epidermidis* and *S. aureus*1 isolates located in the positive region of PCA score plot of PC2 (42%) vs. PC3 (5.95%), showing inter-genus similarities. The band is related to the C-H stretch and combination band region as a result of lipid content in the sample (Osborne, 2000; Genot *et al.*, 2014). In the negative region, three peaks are observed at 1432 nm (protein) 1291 and at 1890 nm (carbohydrates). The separation of the *B. cereus* and *S. epidermidis* isolates are based on these bands. The cell wall of *B. cereus* contains a proteinaceous crystalline arrangement namely the S-layer (Navarre and Schneewind, 1999). The S-layer contains glycan, which consist of disaccharide (carbohydrate) repeating units and could be the reason for the peaks observed (Seltmann & Holst, 2002). In *S. epidermidis* surface proteins such as microbial surface components recognizing adhesive matrix molecules (MSCRAMMs) is observed which is required for the non-covalent bonding of teichoic acids (Otto, 2009). Lastly, when closely examining the associated PCA score plot of PC2 vs. PC3 (Figure 4.22(iii)) along with loading 3 (Figure 4.23(iii)), the *B. cereus* and *S. aureus*1 isolates were strongly associated with the positive region. The prominent bands in the positive region were at 1345 (lipid) and 1896 nm (carbohydrate). In the negative region the bands at 1416 nm (protein) and 1847 nm are as a result of the presence of the *S. epidermidis* isolate. The *S. aureus*2 isolate is also observed in this region, which confirms the inter-genera similarities of the *Staphylococci* samples. The *S. epidermidis* isolate is a coagulase - negative organism that contains a fibrinogen protein, almost identical to the coagulase - positive *S. aureus* species, that is composed of alternating serine and aspartate residues (Navarre and Schneewind, 1999). The absorption of moisture at 1939 nm was associated with the *S. epidermidis* isolate as a result of water content. Again, when closely examining the digital images (**ADDENDUM B**, Figure B21), this sample only contained a thin layer of bacteria on the agar surface. The ratio of powdered growth media to water required to make up liquid culture media is approximately 1:3 and so this observation could be as a result of more of the growth media detected by the instrument than the actual bacteria.

Overall, the plating techniques do not seem to have a too large effect on separating the individual organisms. Considering that all the isolates are Gram-positive, overlapping is inevitable as the vast majority of the cell walls are quite similar. However, despite the known similar chemical compositions of Gram-positive bacteria using either streak or spread plates show potential in separating the organisms. However, it is unclear whether the separation is on account of the technique or the chemical composition of each bacteria. Evaluating the score plots (**Addendum B,** Figures B8 & B11) along with the loading lines (**Addendum B,** Figures B9 & B12) of

Group 6 and **7** is required to compare the separation of the Gram-positive bacteria using both plating techniques. The overall differences can be seen in the spectra (**Addendum B**, Figure B3).

Group 9: PCA results of Gram-negative bacteria – *spread*

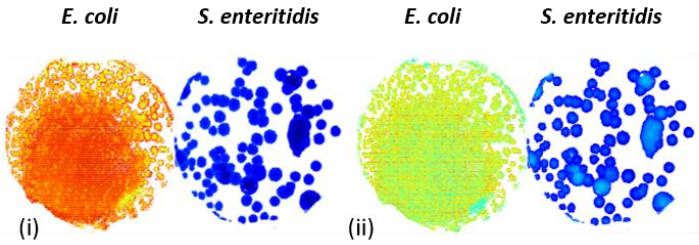

Figure 4.24 Score images of Gram-negative bacteria (i) before and (ii) after pre-processing is applied. SNV produced the best results for Group 9.

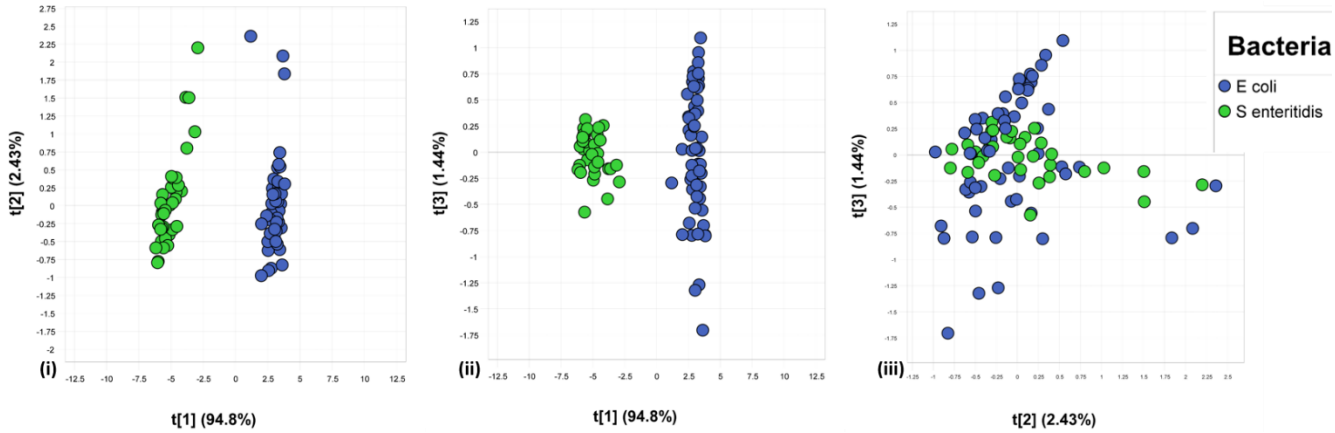

Figure 4.25 PCA score plots illustrating separation of the Gram-negative isolates pre-processed with SNV. The PCA score plot (i), of PC1 (94.8%) vs. PC2 (2.43%) showing clear separation of the bacterial isolates with densely packed data points. For PCA score plot (ii), of PC1 (94.8%) vs. PC3 (1.44%), a similar observation is made, with some spaced out data points. For PCA score plot (iii), PC2 (2.43%) vs. PC3 (1.44%), all the Gram-negative isolates shows overlapping for both plating techniques – no separation.

In **Group 9**, SNV was the preferred pre-processing method. It is evident that the spread plate of *E. coli* suffered from overgrowth as a minimal number of isolated colonies are observed in the score image (Figure 4.24 (i)). The *S. enteritidis* spread plate contained more isolated colonies and a slightly convexed shape with irregular edges. After applying SNV, clear separation was achieved as two distinct clusters are observed (Figure 4.25). The separation seen in PCA score plot (i) of PC1 (94.8%) vs. PC2 (1.44%), could be on account of the dissimilarities contained in the chemical composition of the Gram-negative cell wall. A similar observation was made in PCA

score plot of (ii) PC1 (94.8%) vs. PC3 (2.43%). In the score plot of (iii) PC2 (2.43%) vs. PC3 (1.44%), the isolates show no clear separation.

Distinguishing between Gram-negative bacteria is a more complicated task as the cell wall are believed to be quite similar in composition (Beveridge, 1999). However, using different plating techniques does not have a major effect on separating the organisms. The spread plates of *E. coli* suffered from overgrowth and only a thin layer of bacteria was observed on the surface of the agar, hindering the separation. This in comparison to the *S. enteritidis* samples could have an effect on the separation, as the colonies of these organisms were isolated and thicker in shape. The overlapping data points could be ascribed to the perceived similarities of the Gram-negative cell wall. Both streak (**Addendum B**, Figure B14) and spread plates (Figure 4.25) showed good separation between the organisms. Examining the individual spectra and loadings are required to directly relate the separation to the plating technique (**Addendum B**, Figure B4 – B6).

Group 11 PCA results of Gram-positive and Gram-negative bacteria – *streak and spread*

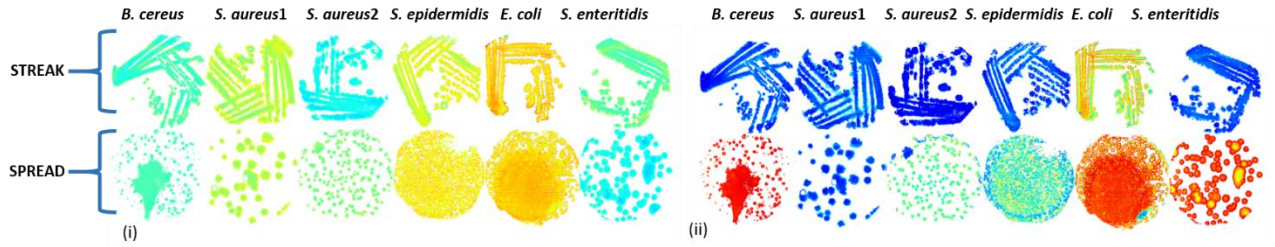

Figure 4.26 Score images of Gram-positive and Gram-negative bacteria for both plating techniques (i) before and (ii) after pre-processing is applied. SNV produced the best results.

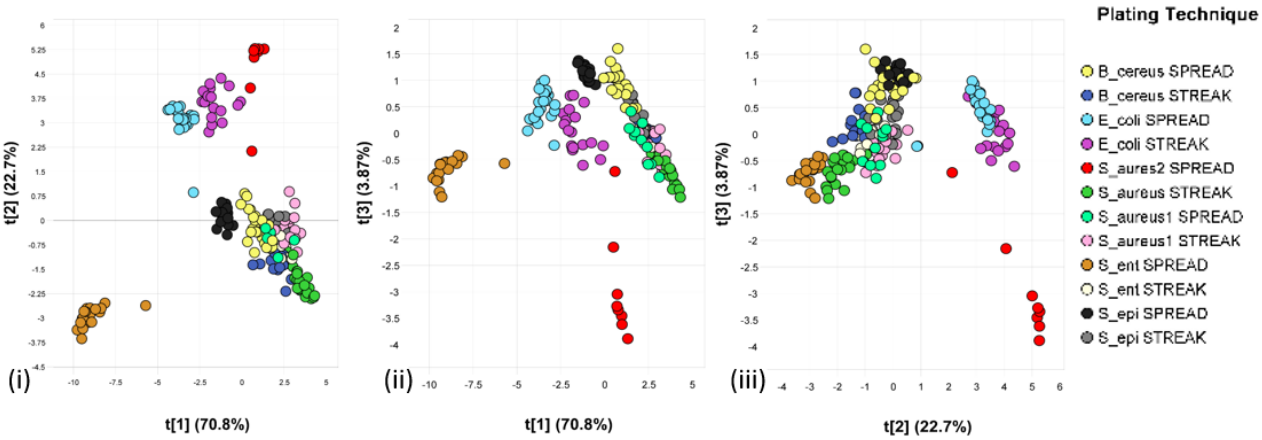

Figure 4.27 PCA score plots (SNV treated) of all six bacterial isolates classed based on plating technique. The PCA score plot (i), of PC1 (70.8%) vs. PC2 (22.7%) showed some isolated clusters of streak and spread plates. In PCA

score plot (ii), of PC1 (70.8%) vs. PC3 (3.87%), a similar observation is made. In PCA score plot (iii), PC2 (22.7%) vs. PC3 (3.87%), overlapping of the isolates using the two different techniques.

Overall comparison

The PCA score plots of **Group 11** were grouped according to the plating techniques used for each individual bacteria. In PCA score plot of PC1 (70.8%) vs. PC2 (22.7%) (Figure 4.27 (i)), the spread plates of the *Salmonella* (orange) isolate formed an isolated cluster in the negative region. In the positive region both streak and spread plates of *S. epidermidis* (grey and black), *B. cereus* (blue and yellow) and *S. aureus*1 (pink and turquoise) clustered together. The clustering observed is on account of similar characteristics in the Gram-positive cell wall. The streak plates of *S. aureus*2 (green) however, is also observed in this region, again confirming similarities in biochemical composition of the bacteria. The streak and spread plates of the *E. coli* isolate (purple and light blue), clustered together. However, the *S. aureus*2 (red) isolate was also observed where the *E. coli* isolates were, therefore the model was not able to clearly separate these two organisms.

When examining PCA score plot of PC1 (70.8%) vs. PC3 (3.87%) (Figure 4.27(ii)), the *S. aureus*2 isolate is now clearly separated from the rest of the isolates and located in the positive region. The separation is therefore not on account of the plating technique but rather the chemical composition of the bacteria. The clustering for the rest of the isolates remained the same as the previous observation. Lastly, when examining PCA score plot of PC2 (22.7%) vs. PC3 (3.87%), overlapping of the Gram-positive (*B. cereus, S. aureus, S. epidermidis*) and Gram-negative (*Salmonella*) bacteria is observed for both plating techniques. The streak and spread plates of *E. coli* formed an isolated cluster and could therefore be separated.

When comparing the PCA results to that of the score images (Figure 4.26) and digital images (**Addendum B**, Figure B21), one reason for the overlapping observed could be as a result of the size and shapes of the colonies. The spread plate of *S. aureus*2 contained smaller colonies, compared to the rest of the Gram-positive isolates.

It could therefore be that the ratio of the pixels belonging to the agar is higher than that of the pixel of the bacterial colonies. The result is that a mixture of the bacteria and agar produce 'impure' spectral results causing the separation. The separation observed for *E. coli* is as a result of the chemical composition that is very different from the rest of the isolates (Panawala, 2017). Lastly, the *S. enteritidis* isolate indicates a low score value before and after being pre-treated which is in agreement with the rest of the streak plates (except *E. coli*) and could therefore be the reason for the separation. It seems that the streak plates show more correlation to one another, which is less prevalent in the spread plates. It is unclear whether it could be that the streak plate method is optimal for bacterial identification with NIR hyperspectral imaging. However, it should be noted that some of the spread plates suffered from overgrowth due to the 24 - hour incubation period and therefore the

separation (or lack thereof) cannot be ascribed to the plating technique. To confirm these observations, the spectra (Figure 4.20) and loadings (**Addendum B**, Figure B10) of the individual bacteria need to be examined.

4.2.3 Partial Least Squares Discriminant Analysis (PLS-DA)

As mentioned previously, three different pre-processing techniques were evaluated to potentially improve the separation and determine which would obtain optimal classification accuracies. The table below depicts the overall results obtained after implementing the different pre-processing treatments. The highlighted values represent the model that obtained the best results for each group.

Table 4. 4 Overview of different pre-processing methods applied for bacterial discrimination

	Performance	Pre-processing		
Group		**SNV**	**SNV & SG**	**SG**
5	Classification accuracy (%)	99.16	99.42	98.08
	Misclassification rate (%)	0.84	0.58	1.92
6	Classification accuracy (%)	99.24	99.18	99.46
	Misclassification rate (%)	0.76	0.82	0.54
7	Classification accuracy (%)	98.94	97.27	99.63
	Misclassification rate (%)	1.06	2.73	0.37
8	Classification accuracy (%)	100	100	100
	Misclassification rate (%)	0	0	0
9	Classification accuracy (%)	100	100	100
	Misclassification rate (%)	0	0	0
10	Classification accuracy (%)	100	100	100
	Misclassification rate (%)	0	0	0
11	Classification accuracy (%)	99.61	99.52	98.90
	Misclassification rate (%)	0.39	0.48	1.10

SNV= Standard normal variate, SG = Savitzky-Golay,

Optimal model selection based on the best classification accuracy after pre-processing

In Table 4.4, the effect of using different pre-processing techniques can be seen. When applying SNV to the PLS-DA model of **Group 5** the classification accuracy obtained was 99.16%. An improved classification accuracy (99.42%) was observed when applying a combination of SNV and SG producing an increase of 0.26%.

When applying only SG, the classification accuracy dropped with 1.34%. The optimal model selected was therefore the combination (SNV and SG) pre-treated model.

For **Group 6,** the model that produced the best results was the SG pre-treated model, obtaining a classification accuracy of 99.46%. The two remaining models only showed a slight decrease in overall classification accuracy. The models treated with SNV and a combination of SNV and SG, obtained classification accuracies of 99.24 and 99.18% respectively. Overall these models all produced excellent results, but the model treated with only SG was preferred. The classification accuracies obtained for **Group 7** was 98.94, 97.27 and 99.63% respectively. Again, the model treated with SG produced the best result. A slight decrease in classification accuracy of 0.69% (SNV treated) and 2.36% (SNV and SG treated) was observed.

For the Gram-negative organisms, all the models performed exceptionally well. The classification accuracies obtained applying the various pre-processing techniques were 100%. It can therefore be concluded that for these particular models, applying all the pre-processing techniques discussed, the models have a high probability of correctly identifying each of the bacteria. It should however be noted that these models only contained two different bacteria and therefore less variation in the chemical components. The variation for the overall model is evident in the spectra, score plots and loading lines of the individual bacteria (**ADDENDUM B**, Figures B6, B17 and B18). The variation could possibly arise from some differences in minor constituents of the bacterial cells as well as the plating technique used. A comparative summary of the overall performance of the optimal PLS-DA models of **Group 5-11** can be seen in Table 4.5 below.

Table 4.5 The performance measures used to assess the optimal PLS-DA results for the different bacterial isolates.

Mosaic	Optimal number of LV's	Classification Accuracy (%)	Sensitivity (%)	Specificity (%)	Misclassification Rate (%)
5	7	99.16	96.72	99.67	0.84
6	8	99.24	98.48	99.39	0.76
7	9	98.94	95.88	99.57	1.06
8	8	100	100	100	0
9	6	100	100	100	0
10	7	100	100	100	0
11	9	99.61	97.73	100	0.39

Comparing the performance measures of the Gram-positive models, we observe that **Group 6** is the best overall model. The misclassification rate (0.76%) for **Group 6** compared to **Group 5** (0.84%) and **7**(1.06%) indicates the excellent predictive ability of the model. When comparing the sensitivity of each model, a decrease of 1.76% and 2.6% can be seen for **Group 5** and **7**. This indicates that these groups produced a higher number of false positives. The reason for the higher misclassification rate in **Group 7** can be ascribed to the increased variability contained in the samples. The different plating techniques produce different sizes and shapes of the isolated colonies. This can be seen in the score images and the digital images (**Addendum B**, Figure B10 & B21).

All the Gram-negative models, obtained excellent results of 100% sensitivity and specificity, proving that both streak and spread plate techniques can be used to separate the individual bacteria. The results are promising as although all Gram-negative bacteria have similar biochemical attributes, some minor constituents are useful in classifying the bacteria independently. Overall, the statistical performance of the models were excellent, showing that irrespective of the plating technique, the models could be used to distinguish between the individual bacterial isolates. A visual representation of the overall performance of the models can be seen in the prediction maps**,** Figures 4.28, 4.29 and 4.30.

Figure 4.28 Prediction map of **Group 7** illustrating overall model performance of Gram-positive bacteria using streak and spread plate techniques.

Figure 4.29 Prediction map of **Group 10** illustrating overall model performance of Gram-negative bacteria using streak and spread plate techniques.

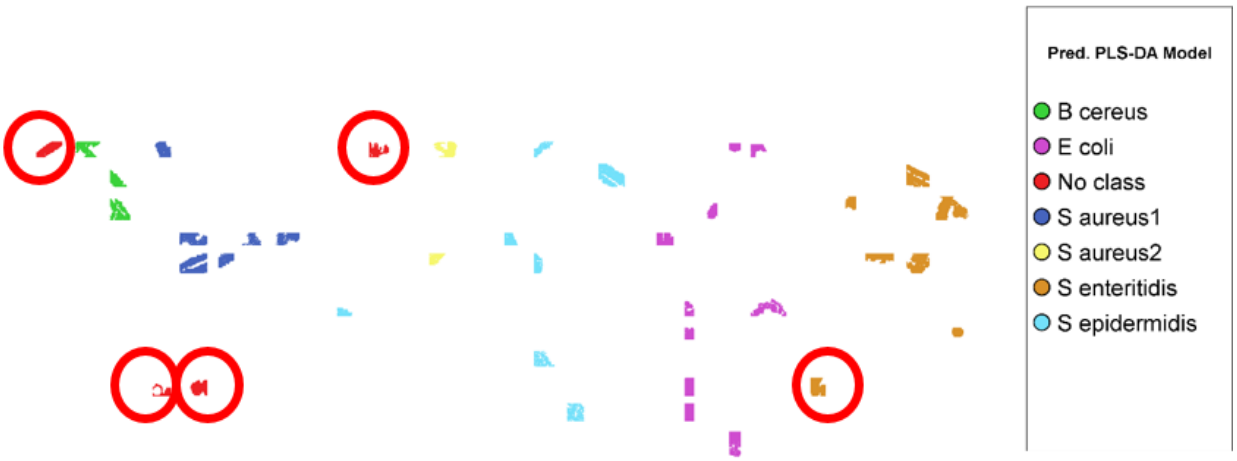

Figure 4.30 Prediction map of **Group 11** illustrating overall model performance of bacteria based on Gram-stain classification, using streak and spread plate techniques. The encircled objects that were not classified correctly include: one object incorrectly classified as *S. epidermidis* (light blue), whereas it was. *S.aureus*2, one object was incorrectly classified as *S. enteritidis* (orange), whereas it was *E. coli* and a total of four objects were not classed at all and assigned to 'no class' (red).

The objects that could not be classified by the model (no-class), are coloured red. One object of the *S.* aureus2 isolate was incorrectly classified as *S. epidermidis* (light blue). Whereas one object of the *E. coli* isolate was incorrectly classified as *S. enteritidis* (orange) as highlighted. Increased variability (bacteria and plating technique) was contained in this model, which could be the reason for a decreased capability to correctly predict all the objects contained in the sample. Overall, the object-wise PLS-DA classification provided excellent accuracy when predicting unknown samples.

In the prediction maps it can be noted that the individual Gram-positive and Gram-negative models performed well as all the objects were correctly classified. PLS-DA classification improved the separation of the individual organisms. The overall Gram-negative model confirms what was observed in the PCA score plots of the same model as a 100% classification accuracy was obtained for this model using PLS-DA classification. Both the streak and spread plate techniques were used in these models and therefore proves that the plating technique does not have an enormous effect on the separation of microorganisms. When examining the digital images (**ADDENDUM B**, Figure B21), of **Group 7** and **10** the samples that suffered from overgrowth did not have an effect on the classification of the bacteria based on their Gram variability. This observation suggests that the separation observed in the individual models are purely based on the chemical composition of the individual bacteria and not on the plating technique.

The model containing all the bacteria using both plating techniques (**Group 11**) however, highlighted the variation observed in the size and shapes of the colonies. The plates suffering from overgrowth was spread plates

of the *B. cereus* and *S. epidermidis* isolates. The sizes of the colonies observed for *S. aureus*2 was larger than that obtained for *S. aureus*1. The objects that were not classed belonged to these samples and the thickness and sizes of the colonies therefore hindered the prediction capabilities of the model. The chemical components, especially the peptidoglycan layer of *E. coli* and *S. enteritidis* are quite similar with some minor differences (Schleifer and Kandler, 1972), which could be the reason for the misclassification.

The observations made in the overall predictions are in correlation with Jiang *et al.*, (2004) that used infrared spectroscopy to elucidate functional groups in Gram-positive and Gram-negative bacteria. In the study it is reported that the isolation process (plating technique) greatly alters the chemical environment of the bacterial cell wall. This suggests that the method used to present the sample to the media does affect the separation of microorganisms. However, the results presented here confirms that an object-wise PLS-DA classification can be used to accurately predict and identify bacteria as the plating technique does not largely affect the identification of each individual organism.

4.3 CONCLUSION

NIR-hyperspectral imaging, in conjunction with MDA could accurately distinguish between bacteria based on Gram-stain reaction as well as pathogenicity. Furthermore, not only was separation possible within the same genera, but among different strains of the same species as well. Both a pixel and object-wise approach was used to determine which would provide optimal results for the accurate identification of bacteria. The results showed that an object-wise approach is superior to that of the pixel-wise approach, as poor separation of the organisms was obtained for the latter. Varying combinations of growth media and plating techniques were used to observe the separation of Gram-positive and Gram-negative bacteria. PLS-DA classification provided excellent overall performance of the models. The models used to evaluate the effect of growth media showed that using different media had no effect on the NIR spectra of microorganisms. The TSA model produced a classification accuracy of 100%. However, it is still unclear whether this indicates that TSA is the better growth media of the two for bacterial classification. The other models, although obtaining a lower classification accuracy percentage, performed equally as well. This proves that the object-wise approach for foodborne pathogenic bacteria has great potential for classification purposes regardless of the growth media used. For the models used to evaluate the effect of different plating techniques, different pre-processing methods were employed. The models produced excellent classification accuracies ranging from 99.16-100%. Although some models had lower classification accuracies, good separation between the bacterial isolates were still observed. The Gram-negative models, only pre-treated with SNV, all produced classification accuracies of 100%. Proving that using different plating techniques has no effect on the separation of foodborne pathogenic bacteria. Although some of the samples suffered from overgrowth due to the long incubation period which resulted in less isolated colonies, this did not greatly affect the overall results. Future work should however focus on implementing a time series analysis to observe how early in the growth phase of bacteria NIR-hyperspectral imaging would be able to distinguish between foodborne pathogenic bacteria. This would allow for more rapid results and ensure a more robust model.

4.4 REFERENCES

Beveridge, T.J. (1999). Structures of Gram-negative cell walls and their derived membrane vesicles. *Journal of Bacteriology*, **181**, 4725–4733.

Bezuidenhout, C. (2018). Near Infrared Hyperspectral Imaging : A Rapid Method For The Differentiation Of Maize Ear Rot Pathogens On Growth Media. *MSc Thesis*. Stellenbosch University, Stellenbosch.

Brewer, J. H. and Carski, T. J. (1950). Petri-dish cover. *Biological Laboratory*. Baltimore, **2**, 1949–1951.

Buchanan, R.L. & Deroever, C.M. (1993). Limits In Assessing Microbiological Food Safety. *Journal of Food Protection*, **56**, 725–729.

Clevers, J.G.P.W. & Kooistra, L. (2006). Using spectral information at the nir water absorption features to estimate canopy water content and biomass. *ISPRS Archives – Volume XXXVI Part 7*, 6.

Cooper, G. M. (2000). The Cell: A Molecular Approach: [www document] URL http://wwwncbi.nlm.nih.gov/books/NBK9839/.html.

Curran, P.J., Kupiec, J.A. & Smith, G.M. (1997). Remote sensing the biochemical composition of a slash pine canopy. *IEEE Transactions on Geoscience and Remote Sensing*, **35**, 415–420.

Dubois, J., Neil Lewis, E., Fry, F.S. & Calvey, E.M. (2005). Bacterial identification by near-infrared chemical imaging of food-specific cards. *Food Microbiology*, **22**, 577–583.

Eady, M., Setia, G. & Park, B. (2019). Detection of Salmonella from chicken rinsate with visible/near-infrared hyperspectral microscope imaging compared against RT-PCR. *Talanta*, **195**, 313–319.

Esbensen, K.H., Guyot, D., Westad, F. & Houmolle, L.P. (2006). An Introduction to Multivariate Data Analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **169**, 387–389.

Farrés, M., Platikanov, S., Tsakovski, S. & Tauler, R. (2015). Comparison of the variable importance in projection (VIP) and of the selectivity ratio (SR) methods for variable selection and interpretation. *Journal of Chemometrics*, **29**, 528–536.

Feng, Y.-Z. & Sun, D.-W. (2014). "Seeing the Bacteria": Hyperspectral Imaging for Bacterial Prediction and Visualisation on Chicken Meat. *NIR news*, **25**, 4–6.

Geladi, P., Isaksson, H., Lindqvist, L., Wold, S. & Esbensen, K. (1989). Principal component analysis of multivariate images. *Chemometrics and Intelligent Laboratory Systems*, **5**, 209–220.

Genot, V., Bock, L., Dardenne, P. & Colinet, G. (2014). L'intérêt de la spectroscopie proche infrarouge en analyse de terre (synthèse bibliographique). *Biotechnology, Agronomy and Society and Environment*, **18**, 247–261.

Gross, J.E., Goetz, S.J. & Cihlar, J. (2009). Application of remote sensing to parks and protected area monitoring: Introduction to the special issue. *Remote Sensing of Environment*, **113**, 1343–1345.

Havelaar, A.H., Brul, S., Jong, A. de, Jonge, R. de, Zwietering, M.H. & Kuile, B.H. ter. (2010). Future challenges to microbial food safety. *International Journal of Food Microbiology*, **139**, S79–S94.

Jiang, W., Saxena, A., Song, B., Ward, B.B., Beveridge, T.J. & Myneni, S.C.B. (2004). Elucidation of functional groups on Gram-positive and Gram-negative bacterial surfaces using infrared spectroscopy. *Langmuir*, **20**, 11433–11442.

Jury, W.A. & Vaux, H. (2005). The role of science in solving the world's emerging water problems. *Proceedings of the National Academy of Sciences*, **102**, 15715–15720.

Kammies, T. (2018). *The evaluation of foodborne pathogenic bacteria using near infrared (NIR) hyperspectral imaging and multivariate image analysis*.

Kammies, T.L., Manley, M., Gouws, P.A. & Williams, P.J. (2016). Differentiation of foodborne bacteria using NIR hyperspectral imaging and multivariate data analysis. *Applied Microbiology and Biotechnology*, **100**, 9305–9320.

Kamruzzaman, M., ElMasry, G., Sun, D.W. & Allen, P. (2012). Prediction of some quality attributes of lamb meat using near-infrared hyperspectral imaging and multivariate analysis. *Analytica Chimica Acta*, **714**, 57–67.

Kucheryavskiy, S. (2013). A new approach for discrimination of objects on hyperspectral images. *Chemometrics and Intelligent Laboratory Systems*, **120**, 126–135.

Mandal, P.K., Biswas, A.K., Choi, K. & Pal, U.K. (2011). Methods for Rapid Detection of Foodborne Pathogens: An Overview. *American Journal of Food Technology*.

Maquelin, K., Choo-Smith, L.P. in., Vreeswijk, T. Van, Endtz, H.P., Smith, B., Bennett, R., Bruining, H.A. & Puppels, G.J. (2000). Raman spectroscopic method for identification of clinically relevant microorganisms growing on solid culture medium. *Analytical Chemistry*, **72**, 12–19.

Mlynáriková, K., Samek, O., Bernatová, S., Růžička, F., Ježek, J., Hároniková, A., Šiler, M., Zemánek, P. & Holá, V. (2015). Influence of Culture Media on Microbial Fingerprints Using Raman Spectroscopy. *Sensors*, **15**, 29635–29647.

Mosier-Boss, P.A. (2017). Review on SERS of bacteria. *Biosensors*, **7**.

Nakakimura, Y., Vassileva, M., Stoyanchev, T., Nakai, K., Osawa, R., Kawano, J. & Tsenkova, R. (2012). Extracellular metabolites play a dominant role in near-infrared spectroscopic quantification of bacteria at food-safety level concentrations. *Analytical Methods*, **4**, 1389–1394.

Navarre, W.W. & Schneewind, O. (1999). Surface proteins of Gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. *Microbiology and molecular biology reviews : MMBR*, **63**, 174–229.

Osborne, B.G. (2000). Near-Infrared Spectroscopy in Food Analysis. *Encyclopedia of Analytical Chemistry*, 1–14.

Panawala, L. (2017). Difference Between Gram Positive and Gram Negative Bacteria Stunning images of cells Discover how scientists use Main Difference – Gram Positive vs Gram Negative Bacteria. *Pediaa*, 13.

Park, B., Lawrence, K.C., Windham, W.R. & Smith, D.P. (2006). Performance of hyperspectral imaging system for poultry surface fecal contaminant detection. *Journal of Food Engineering*, **75**, 340–348.

Park, B., Seo, Y., Yoon, S.C., Hinton, A., Windham, W.R. & Lawrence, K.C. (2015). Hyperspectral microscope imaging methods to classify Gram-positive and Gram-negative foodborne pathogenic bacteria. *Transactions of the ASABE*, **58**, 5–16.

Ramírez-Castillo, F., Loera-Muro, A., Jacques, M., Garneau, P., Avelar-González, F., Harel, J. & Guerrero-Barrera, A. (2015). Waterborne Pathogens: Detection Methods and Challenges. *Pathogens*, **4**, 307–334.

Rodriguez-Saona, L.E., Khambaty, F.M., Fry, F.S. & Calvey, E.M. (2001). Rapid detection and identification of bacterial strains by Fourier transform near-infrared spectroscopy. *Journal of Agricultural and Food Chemistry*, **49**, 574–579.

Salton, M.R.J., Kim K.S. (2018). Structure. In: Baron S, editor. Medical Microbiology. 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston; 1996. Chapter 2. Available from: https://www.ncbi.nlm.nih.gov/books/NBK8477/

Schleifer, K.H. & Kandler, O. (1972). Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriological reviews*, **36**, 407–477.

Shonhiwa, A.M., Ntshoe, G., Essel, V., Thomas, J. & Mccarthy, K. (2017). A Review of Foodborne Disease Outbreaks Reported To the Outbreak Response Unit. *Nicd*, **16**, 1–6.

Snee, R.D. (1977). Validation of Regression Models: Methods and Examples. *Technometrics*, **19**, 415–428.

Wang, K., Pu, H. & Sun, D.W. (2018a). Emerging Spectroscopic and Spectral Imaging Techniques for the Rapid Detection of Microorganisms: An Overview. *Comprehensive Reviews in Food Science and Food Safety*, **17**, 256–273.

Wang, K., Pu, H. & Sun, D.W. (2018b). Emerging Spectroscopic and Spectral Imaging Techniques for the Rapid Detection of Microorganisms: An Overview. *Comprehensive Reviews in Food Science and Food Safety*, **17**, 256–273.

Williams, P., Kammies, T.-L., Gouws, P. & Manley, M. (2019). Effect of colony age on near infrared hyperspectral images of foodborne bacteria. *Journal of Spectral Imaging*, **1**, 1–12.

Wilson, R.H., Nadeau, K.P., Jaworski, F.B., Tromberg, B.J. & Durkin, A.J. (2015). Review of short-wave infrared spectroscopy and imaging methods for biological tissue characterization. *Journal of Biomedical Optics*, **20**, 30901.

Witkowska, E., Niciński, K., Korsak, D., Szymborski, T. & Kamińska, A. (2019). Sources of variability in SERS spectra of bacteria: comprehensive analysis of interactions between selected bacteria and plasmonic nanostructures. *Analytical and Bioanalytical Chemistry*, **411**, 2001–2017.

World Health & organization. (2015). *No Title*.

Xu, J.L., Esquerre, C. & Sun, D.W. (2018). Methods for performing dimensionality reduction in hyperspectral image classification. *Journal of Near Infrared Spectroscopy*, **26**, 61–75.

CHAPTER 5

General Discussion and Conclusion

Foodborne illnesses are the number one food safety concern globally. Management of microbial food safety issues are becoming problematic as food safety and quality are clearly being compromised (Havelaar *et al.*, 2010). In South Africa alone, 327 foodborne diseases were reported over a five year (2013 – 2017) period (Shonhiwa *et al.*, 2017). This observation was further exacerbated in 2018, as a recent multi-country outbreak of Listeriosis claimed at least 200 lives (WHO, 2018). Fifteen countries were affected including Angola, Botswana, Democratic Republic of the Congo, Ghana, Lesotho, Madagascar, Malawi, Mauritius, Mozambique, Namibia, Nigeria, Swaziland, Uganda, Zambia and Zimbabwe. This could be a result of delays caused in the detection of foodborne pathogenic bacteria.

Microbiological testing plays a vital role in food manufacturing and processing, to ensure food safety. Traditional testing methods are usually based on cultivating bacteria grown from pure cultures and making use of viable counting measurements. These methods include both microbiological and biochemical procedures that provide accurate results and are reliable. In order to obtain presumptive positives however, specific stepwise procedures have to be taken, which could cause a delay in the results. Bacterial cells need to be afforded sufficient time to multiply to visible colonies, which could take up to seven days in certain instances. Highly specialised staff is required to assess the colonies visually, which poses additional challenges as competitive microflora, often confluently grow together and cannot be distinguished by visual observation. This leads to large volumes of products being distributed before being tested, ultimately leading to product recalls and or multiple outbreaks. Improving early detection methods to implement in routine screening, is currently in high demand. It is therefore crucial for the food industry to develop rapid screening techniques that can address some of these time delays associated with the accurate detection of microorganisms.

NIR - hyperspectral imaging can be used as a complimentary, rapid screening tool in conjunction with traditional microbiological methods. It has the ability to provide a spectrum for each pixel in the image (Kamruzzaman *et al.*, 2012). The technique is growing in popularity and is often used to identify foodborne microorganisms with the goal to use it as an initial step for industrial applications (Park *et al.*, 2006, 2015; Gross *et al.*, 2009; Feng and Sun, 2014; Kammies *et al.*, 2016; Bezuidenhout, 2018; Williams *et al.*, 2019). The method presented here addresses the effect different growth media and plating techniques have on the NIR spectra of bacteria. This study found that NIR-HSI could successfully be used in combination with conventional culturing techniques to accurately distinguish between bacteria based on Gram-stain classification as well as

pathogenicity. Ideally, the method should potentially be implemented for food products contaminated with foodborne pathogenic bacteria.

In this study, NIR-HSI coupled with chemometrics was evaluated as a discriminant technique for foodborne pathogenic bacteria cultured on growth media. An important initial step required was to use principal component analysis (PCA), to explore the data and reduce the dimensionality. Both a pixel and object-wise approach was evaluated and compared. Evaluating the two approaches proved that the object-wise approach was preferred for the distinction of foodborne pathogenic bacteria. Since it was evident in the PCA models that distinction was possible, developing classification models, with PLS-DA, were required to determine how the models would perform on future unknown samples. The models were constructed to distinguish between bacteria based on their Gram-stain classification (Gram-positive or Gram-negative), as well as pathogenicity. The object-wise approach achieved classification accuracies of 99.55 – 100%.

Accurate distinction of foodborne pathogenic bacteria on two different growth media is indeed possible. Although the growth media had a slight effect on the intensity of the spectra, it did not greatly influence the separation of the individual bacteria. The plating techniques do not affect the ability of the model to identify foodborne pathogens as classification accuracies of 99.42 – 100% (streak) and 99.46 – 100 (spread) were obtained. The model used to distinguish pathogenic bacteria from (*S. aureus*) non-pathogenic (*S. epidermidis*) *bacteria* obtained a classification accuracy of 98.72%. This confirms that NIR-HSI and chemometrics, using an object-wise classification, can accurately distinguish between foodborne pathogenic bacteria within the same genera as well as different strains.

To further optimize the classification models, different pre-processing techniques were evaluated. The models for the Gram-negative bacteria (pre-treated with SNV), all produced classification accuracies of 100%. Compared to the study by Kammies, *et al*., (2016), where poor separation of *E. coli* (53.9%) and *Salmonella* (2.34%) was obtained, the current study proved that Gram-negative bacteria could be distinguished using an object-wise approach to improve the classification accuracies. Throughout this study, we observed that the *E. coli* isolate differed from the rest of the isolates, and this was ascribed to the unique characteristic of the peptidoglycan layer contained in the bacterial cell wall as per Salton & Kim, (1996), which could be the reason for the excellent separation. The models containing Gram-positive bacteria obtained excellent accuracies of 99.42-99.63%. Various statistical parameters such as the sensitivity, specificity and misclassification rate need to be considered as it influences the classification accuracy. The sensitivity of the models ranged from 95.88 – 100%, whereas the specificity of the models was all above 96.64%. The models presented here obtained excellent classification accuracies ranging from 98.29 – 100% and could all be implemented as a rapid screening tool to use in a food microbiology laboratory where swift, real-time results are required.

NIR-HSI is a candidate technology that provides a spectral 'map' of biochemical patterns of foodborne microorganisms. With this in mind, the results can be used to construct spectral data libraries that would be readily available for food analysis. In order to implement the model in the laboratory, more foodborne pathogenic bacteria (e.g. *Campylobacter, Clostridium botulinum E. coli 0157* (STEC) and *Listeria monocytogenes*) should be included as well as more growth media (e.g. All culture agar (ACA); Columbia blood agar; brain heart infusion (BHI) agar; Mueller Hinton Agar (2 M-H 2 Agar)). The method should also be validated using plating techniques different to the ones described here (e.g. spot plates). Furthermore, it was evident that some of the samples suffered from overgrowth due to the 24 h incubation period. A time-series study should be conducted to determine which bacterial growth curve phase produces optimal classification results. Eady and co-workers (2019), used hyperspectral microscope imaging (HMI) to evaluate incubation times for five different *Salmonella* serotypes and found an 8 h incubation time to be sufficient (Eady *et al.*, 2019). This information is beneficial as it implies that the time spent on cultivating bacteria could further be reduced and used in combination with NIR-HSI to produce results that would be of great value for the food industry. To further optimize the method and ensure robustness, different classification algorithms such as hierarchical cluster analysis (HCA), linear discriminant analysis (LDA) or soft independent model of class analogy (SIMCA), should be explored.

NIR hyperspectral imaging can be used in conjunction with traditional microbiology to serve as a rapid screening tool to distinguish between foodborne pathogenic bacteria within the same genus and species. This provides fast and accurate results that can be implemented in the food industry to reduce the distribution of large volumes of contaminated products reaching the consumer. This could ultimately reduce foodborne outbreaks and therefore greatly contribute to food safety.

5.1 REFERENCES

Bezuidenhout, C. (2018). *Near Infrared Hyperspectral Imaging : A Rapid Method For The Differentiation Of Maize Ear Rot Pathogens On Growth Media*.

Eady, M., Setia, G. & Park, B. (2019). Detection of Salmonella from chicken rinsate with visible/near-infrared hyperspectral microscope imaging compared against RT-PCR. *Talanta*, **195**, 313–319.

Feng, Y.-Z. & Sun, D.-W. (2014). "Seeing the Bacteria": Hyperspectral Imaging for Bacterial Prediction and Visualisation on Chicken Meat. *NIR news*, **25**, 4–6.

Gross, J.E., Goetz, S.J. & Cihlar, J. (2009). Application of remote sensing to parks and protected area monitoring: Introduction to the special issue. *Remote Sensing of Environment*, **113**, 1343–1345.

Havelaar, A.H., Brul, S., Jong, A. de, Jonge, R. de, Zwietering, M.H. & Kuile, B.H. ter. (2010). Future challenges to

Kammies, T. (2018). *The evaluation of foodborne pathogenic bacteria using near infrared (NIR) hyperspectral imaging and multivariate image analysis*.

Kammies, T.L., Manley, M., Gouws, P.A. & Williams, P.J. (2016). Differentiation of foodborne bacteria using NIR hyperspectral imaging and multivariate data analysis. *Applied Microbiology and Biotechnology*, **100**, 9305–9320.

Kamruzzaman, M., ElMasry, G., Sun, D.W. & Allen, P. (2012). Prediction of some quality attributes of lamb meat using near-infrared hyperspectral imaging and multivariate analysis. *Analytica Chimica Acta*, **714**, 57–67.

Park, B., Lawrence, K.C., Windham, W.R. & Smith, D.P. (2006). Performance of hyperspectral imaging system for poultry surface fecal contaminant detection. *Journal of Food Engineering*, **75**, 340–348.

Park, B., Seo, Y., Yoon, S.C., Hinton, A., Windham, W.R. & Lawrence, K.C. (2015). Hyperspectral microscope imaging methods to classify Gram-positive and Gram-negative foodborne pathogenic bacteria. *Transactions of the ASABE*, **58**, 5–16.

Salton, M.R.J., Kim K.S. (2018). Structure. In: Baron S, editor. Medical Microbiology. 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston; 1996. Chapter 2. Available from: https://www.ncbi.nlm.nih.gov/books/NBK8477/

Shonhiwa, A.M., Ntshoe, G., Essel, V., Thomas, J. & Mccarthy, K. (2017). A Review of Foodborne Disease Outbreaks Reported To the Outbreak Response Unit. *Nicd*, **16**, 1–6.

Williams, P., Kammies, T.-L., Gouws, P. & Manley, M. (2019). Effect of colony age on near infrared hyperspectral images of foodborne bacteria. *Journal of Spectral Imaging*, **1**, 1–12.

ADDENDUM A
Supplementary information pertaining to Chapter 4.1

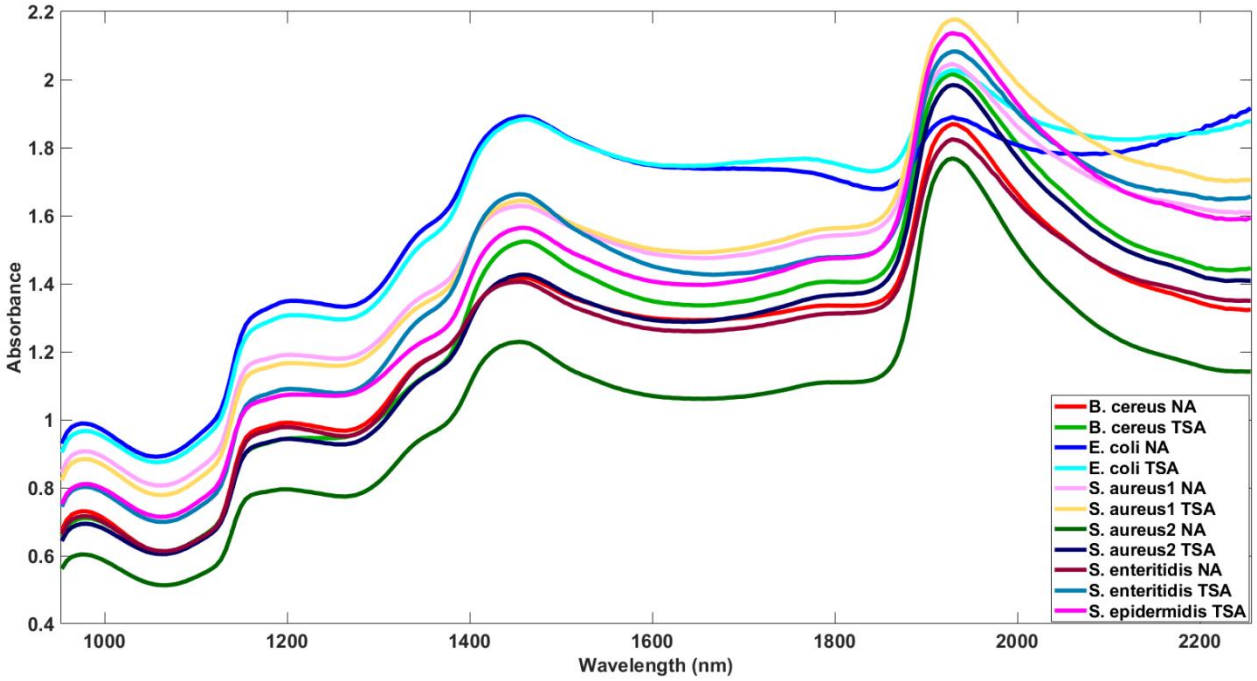

Figure A1 Average unprocessed spectra obtained for individual bacterial isolates on nutrient agar (NA) and tryptic soy agar (TSA).

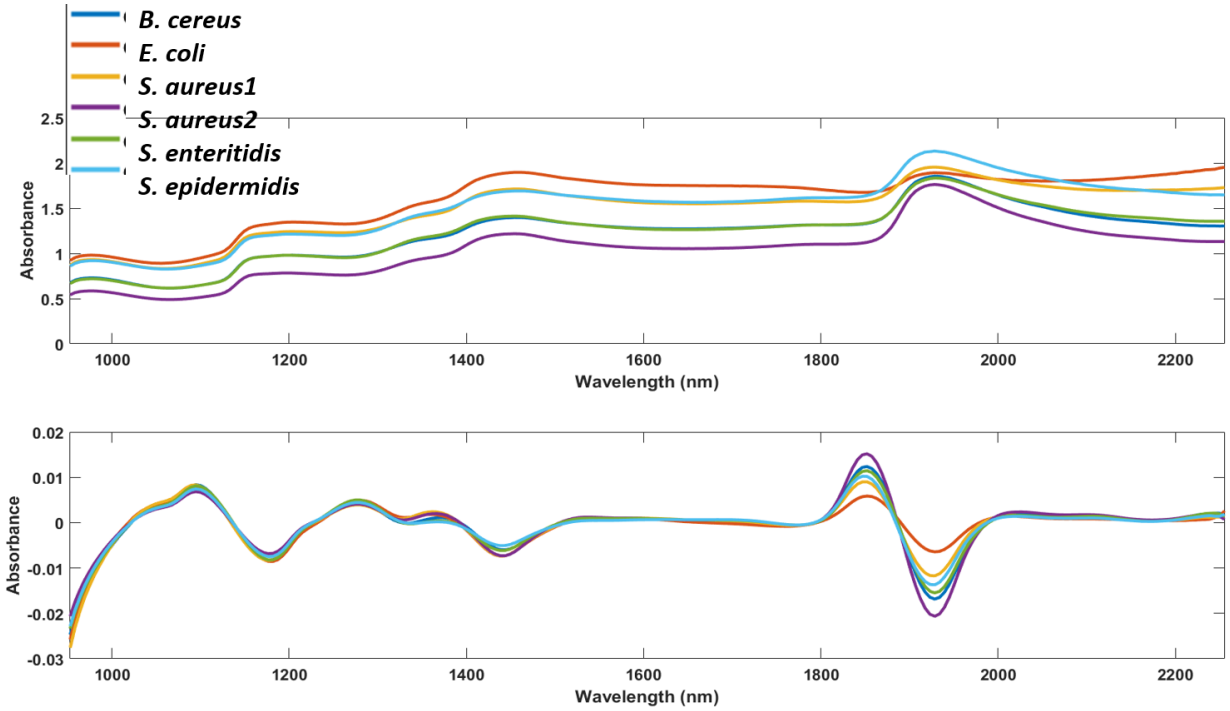

Figure A2 Average unprocessed and derivative spectra obtained for bacterial isolates on nutrient agar (NA)

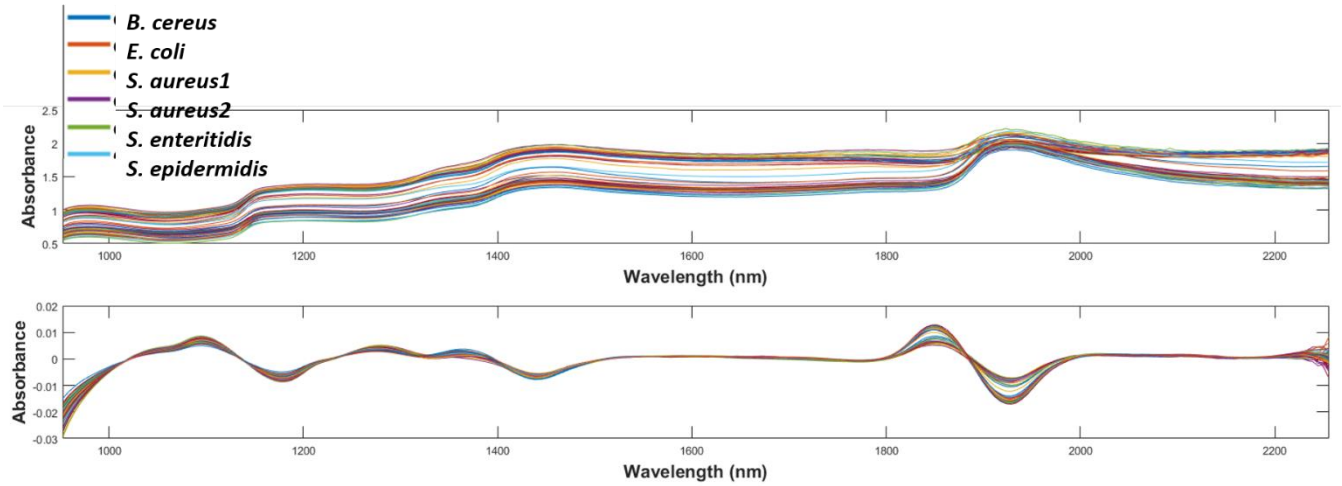

Figure A3 Average unprocessed and derivative spectra obtained for bacterial isolates on tryptic soy agar (TSA).

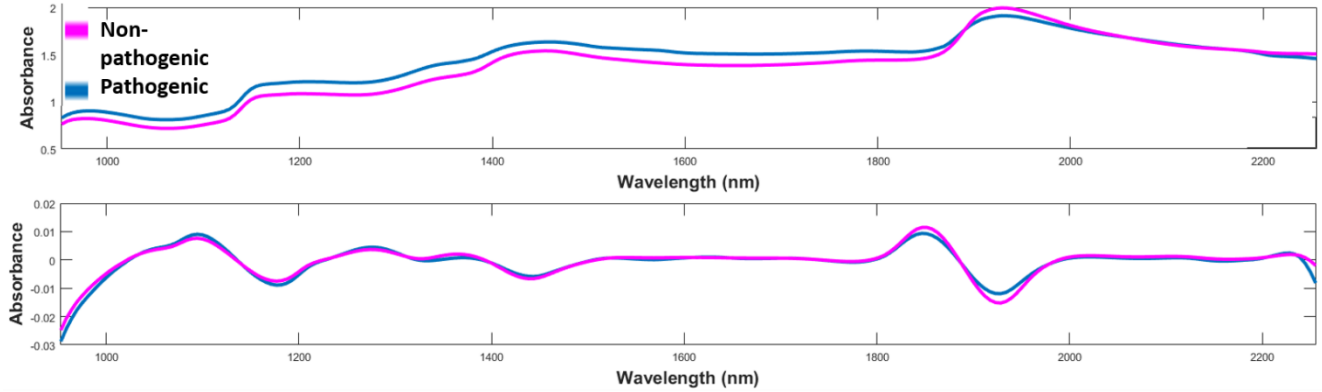

Figure A4 Differences in average unprocessed and derivative spectra obtained for pathogenic and non-pathogenic foodborne pathogens on NA and TSA

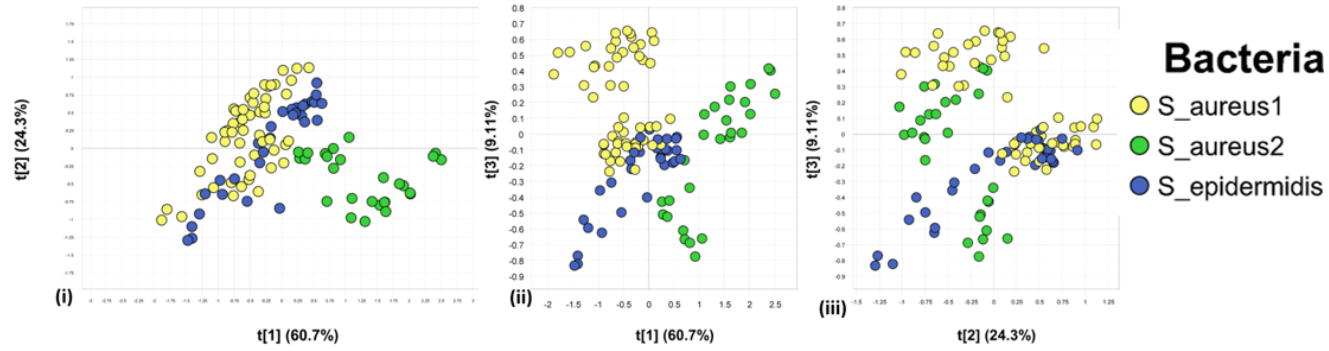

Figure A5 Object-wise mosaic illustrating separation between the pathogenic vs. non-pathogenic bacteria as per the individual organisms. The PCA score plot (i), of PC1 (60.7%) vs. PC2 (24.3%) clustering observed for some of the pathogenic bacteria – possible distinction. PCA score plot (ii), in the direction of PC1 (60.7%) vs. PC3 (9.11%), clustering of both pathogenic isolates (*S. aureus* 1 & 2). In PCA score plot (iii), PC2 (24.3%) vs. PC3 (9.11%) – similar observation.

Digital Images

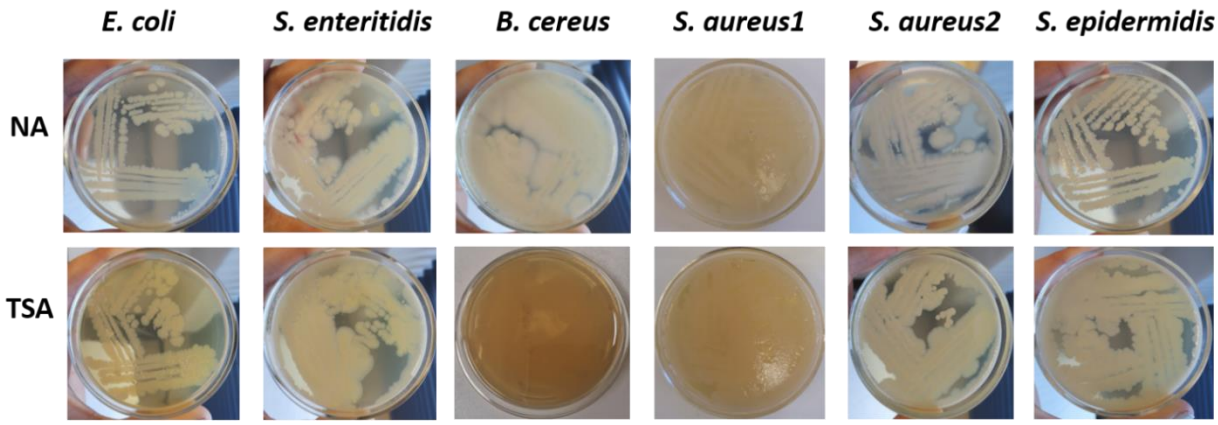

Figure A6 Digital images of foodborne pathogenic bacteria on different growth media (NA and TSA) used in the study. Overgrowth evident on NA: *S. enteritidis*, *B. cereus* and *S. aureus*2; TSA: *S. enteritidis*, *S. aureus*1, *S. aureus*2 and *S. epidermidis*.

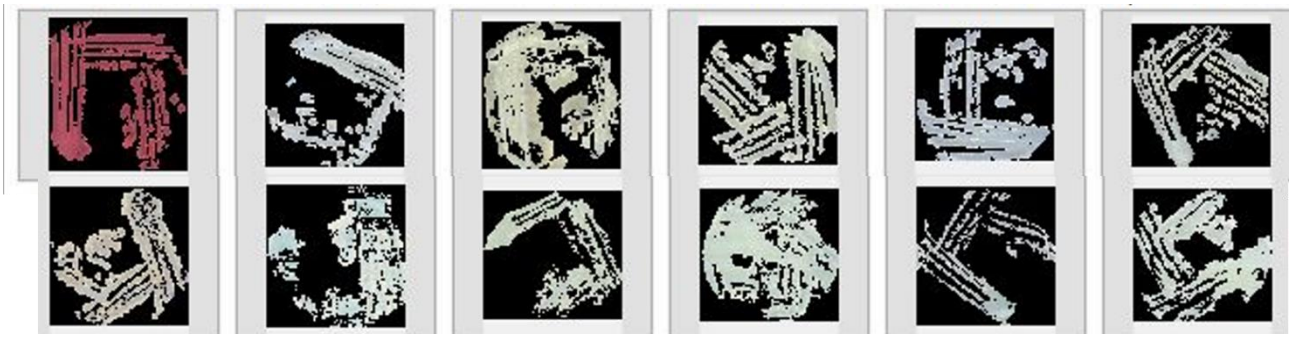

Figure A7 Accompanying RGB images of foodborne pathogenic bacteria on different growth media (NA and TSA) after image cleaning. Care was taken to remove most of the overgrowth that occurred. However, some clumps remained. Thickness of colonies also illustrated Figure A7.

ADDENDUM B
Supplementary information pertaining to Chapter 4.2

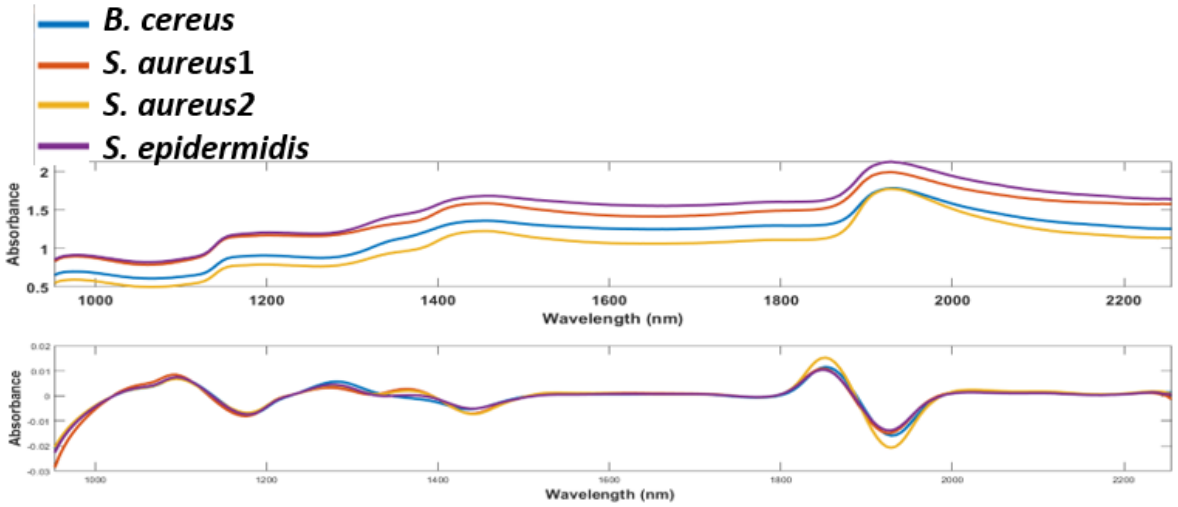

Figure B 1 Average unprocessed and derivative spectra obtained for Gram-positive bacterial isolates – streak plates.

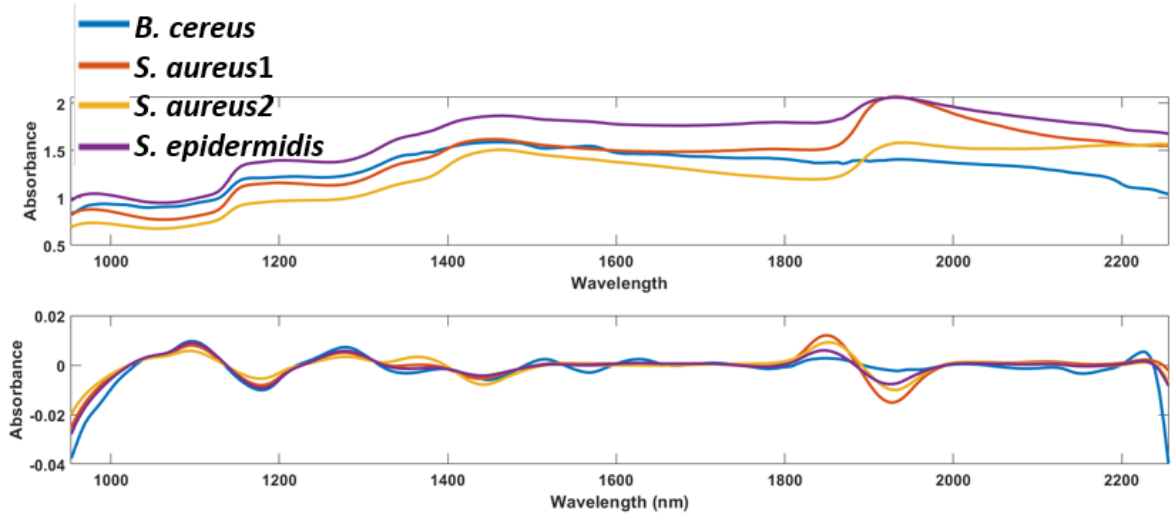

Figure B2 Average unprocessed and derivative spectra obtained for Gram-positive bacterial isolates – spread plates.

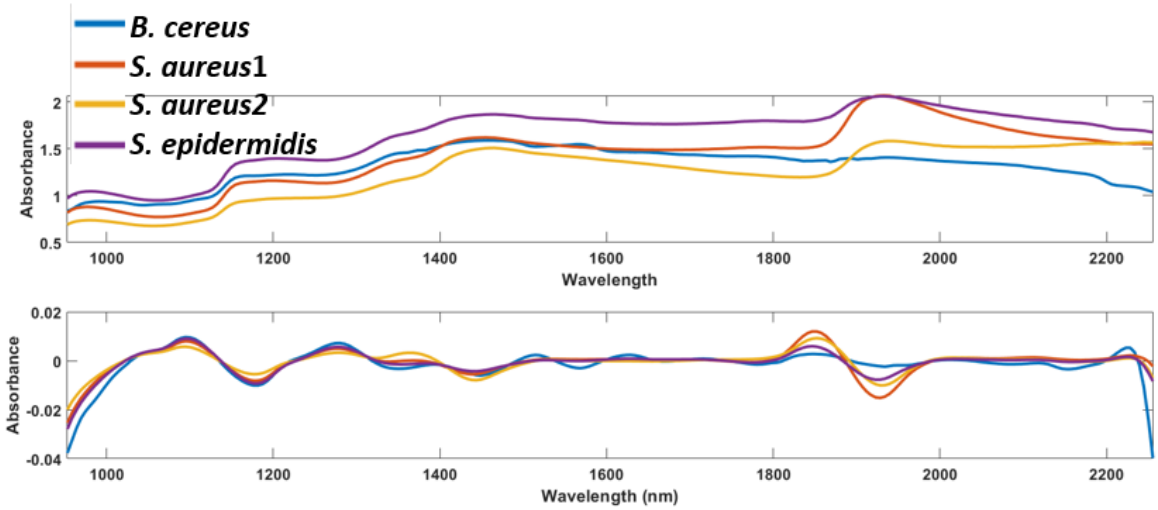

Figure B3 Average unprocessed and derivative spectra obtained for Gram-positive bacterial isolates – streak plates & spread plates.

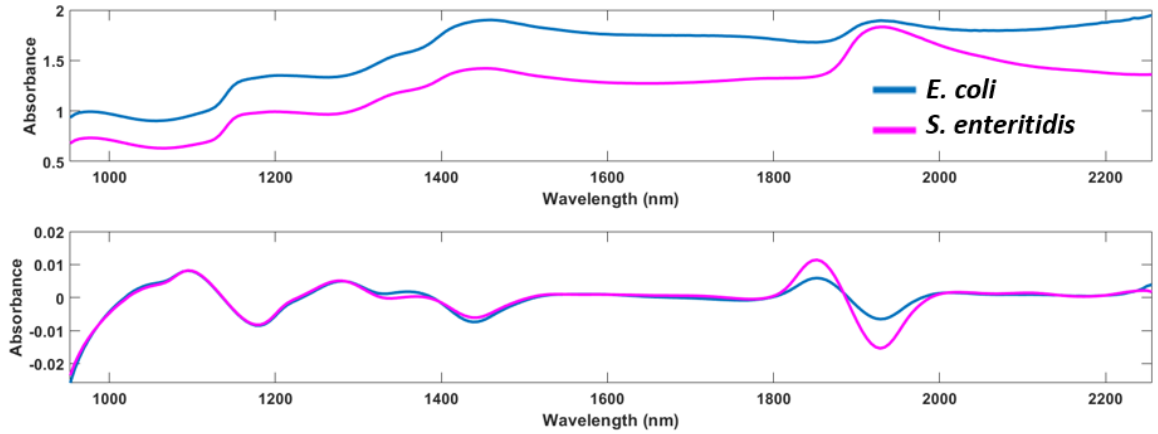

Figure B4 Average unprocessed and derivative spectra obtained for Gram-negative bacterial isolates – streak plates.

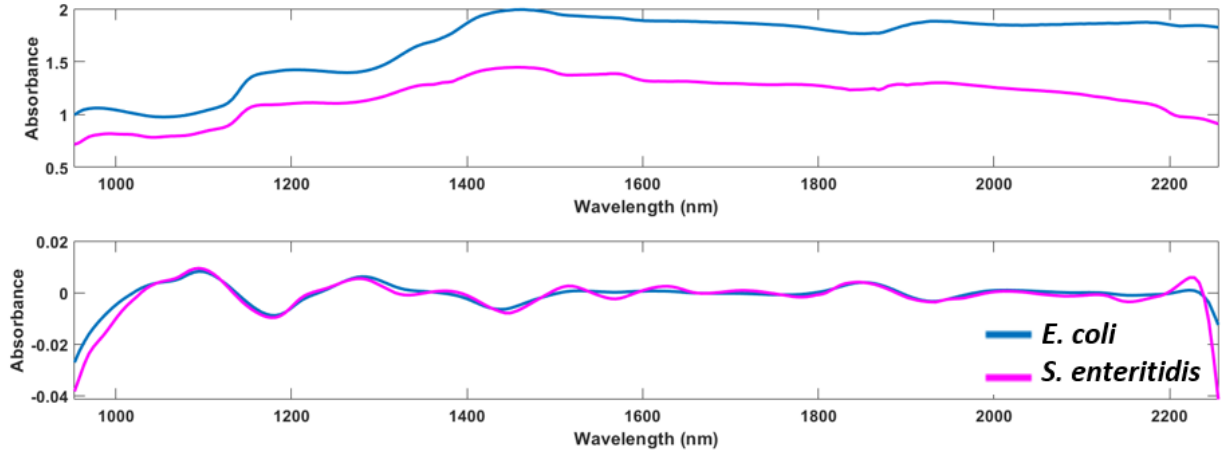

Figure B5 Average unprocessed and derivative spectra obtained for Gram-negative bacterial isolates – spread plates.

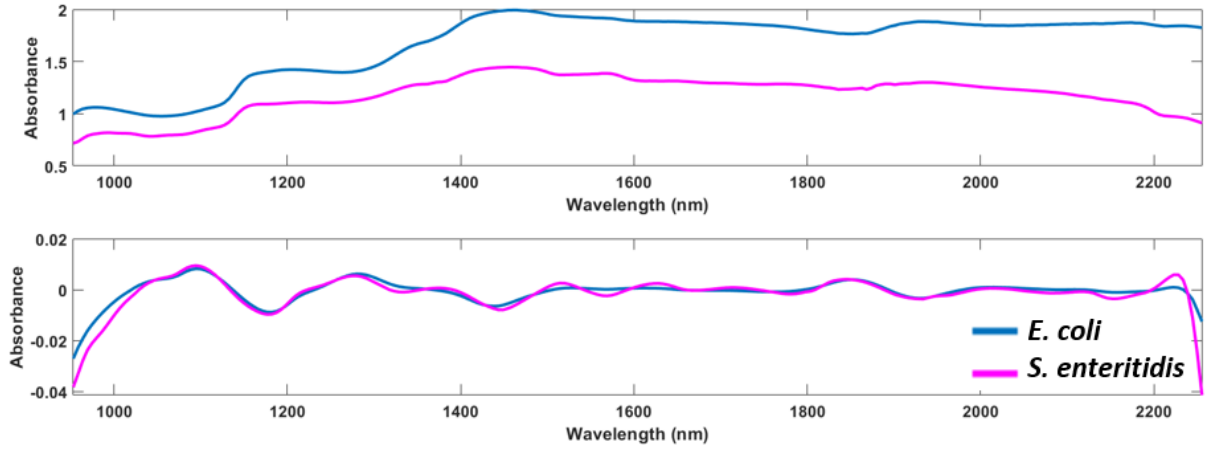

Figure B6 Average unprocessed and derivative spectra obtained for Gram-negative bacterial isolates – streak & spread plates.

Group 6: PCA results of Gram-positive bacteria – *spread*

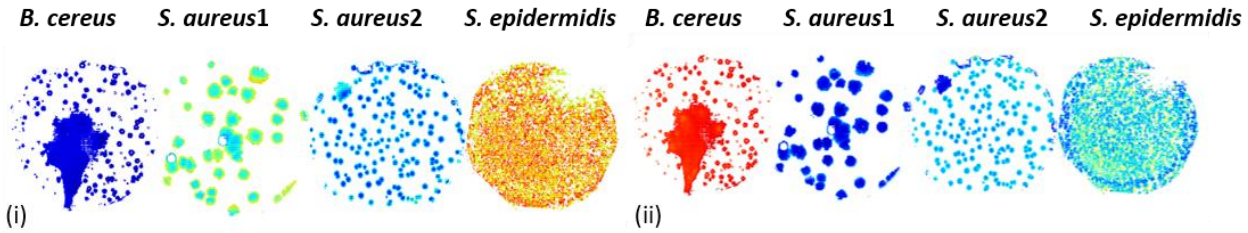

Figure B7 Illustration of score images (i) before and (ii) after pre-processing is applied. SG was the optimal pre-processing method for **Group 6**.

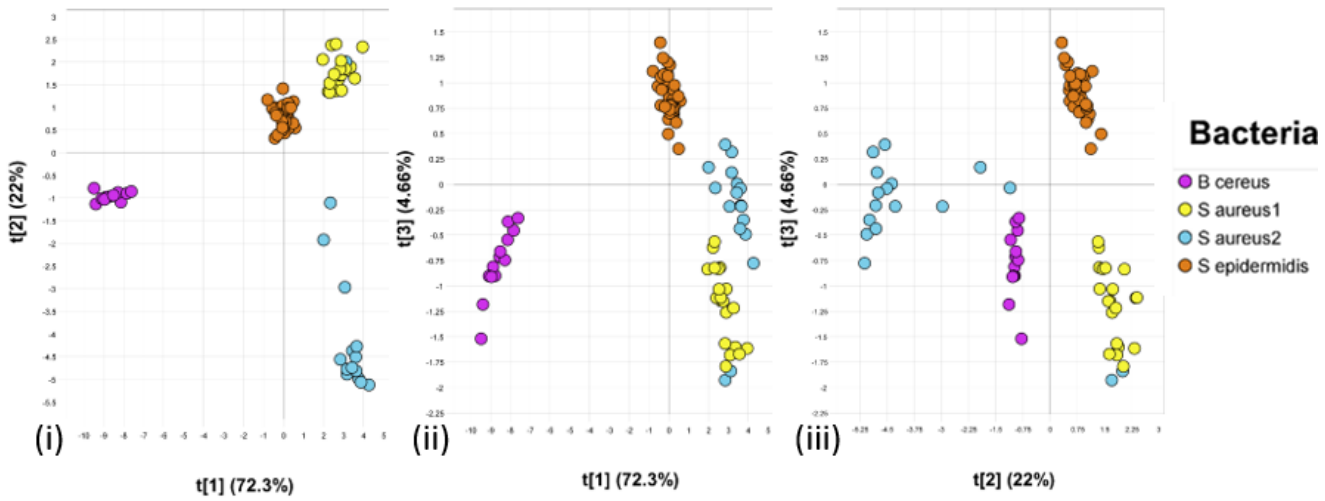

Figure B8 PCA score plots illustrating separation of the Gram-positive isolates treated with SG. The PCA score plot (i), of PC1 (72.3%) vs. PC2 (22%) clear separation of the bacterial isolates. The PCA score plot (ii), of PC1 (72.3%) vs. PC3 (4.66%), some overlapping observed for *S*. *aureus*1 and *S*. *aureus*2. For PCA score plot (iii), PC2 (22%) vs. PC3 (4.66%), a similar observation is made.

Loadings Group 6

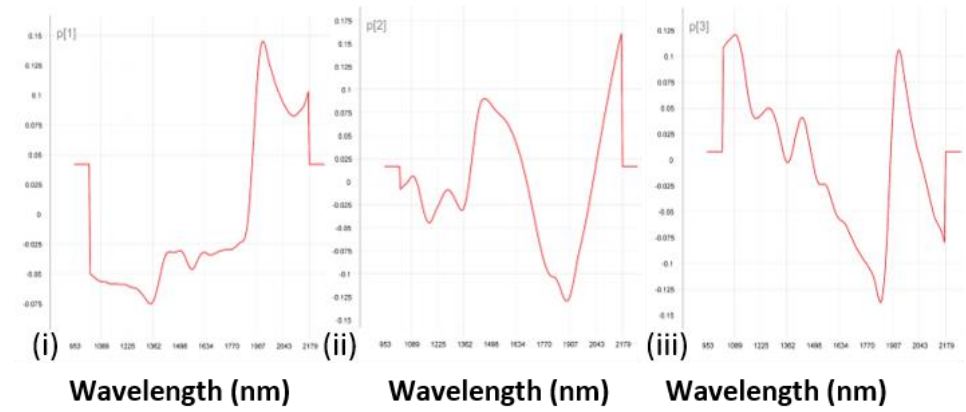

Figure B9 PCA loading lines of **Group 6**. In (i) PC1 a peak is observed in the positive region at 1934 nm and in the negative region at 2179 nm. For (ii) PC2, a sharp peak is observed at 2179 nm (inversed in PC1) and in the negative region, we observe a band at 1907 nm. Lastly, for (iii) PC3 at 1934 nm (pos. region) and at 1847 and 2179 nm (neg. region).

Group 7: PCA results of Gram-positive bacteria – *streak and spread*

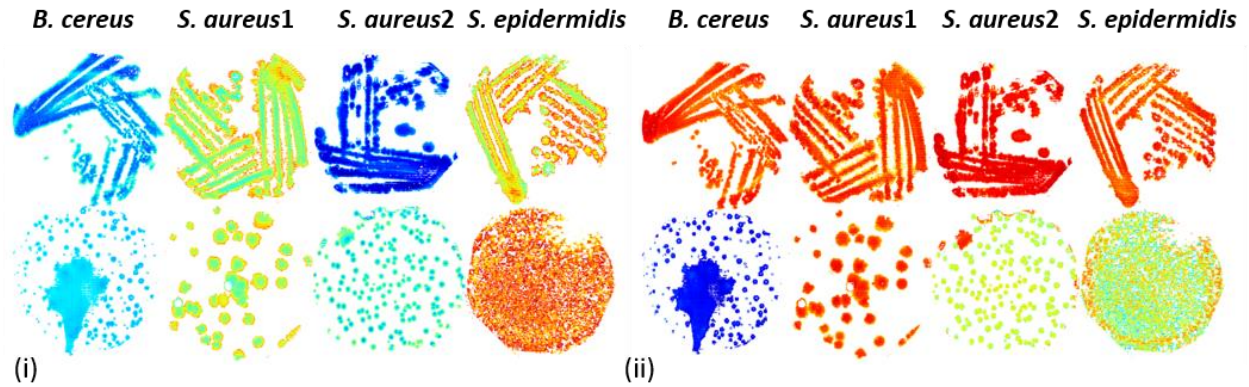

Figure B10 Illustration of score images (i) before and (ii) after pre-processing is applied. Savitzky-Golay (d_2; 5pt.) provided the best separation **Group 7**.

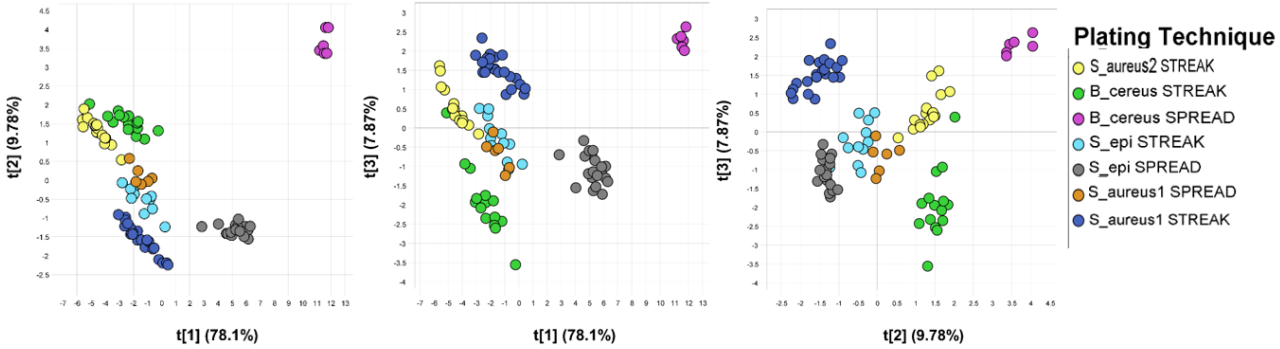

Figure B11 PCA score plots illustrating separation of the Gram-positive isolates on streak and spread plates treated with SGd_2; (5pt.). In PCA score plot (i), in PC1 (78.1%) vs. PC2 (9.78%) the spread plates of *B*. *cereus*, *S*. *aureus*1 and *S*. *epidermidis*, formed densely packed clusters. In the PCA score plot (ii), in PC1 (78.1%) vs. PC3 (7.87%), some overlapping observed for the spread plates of *S*. *aureus*1 and *S*. *aureus*2 and *S*. *epidermidis*. For PCA score plot (iii), PC2 (9.78%) vs. PC3 (7.87%), separation is observed for all the isolates, (except *S*. *aureus* 1) on account of the plating technique.

Loadings Group 7

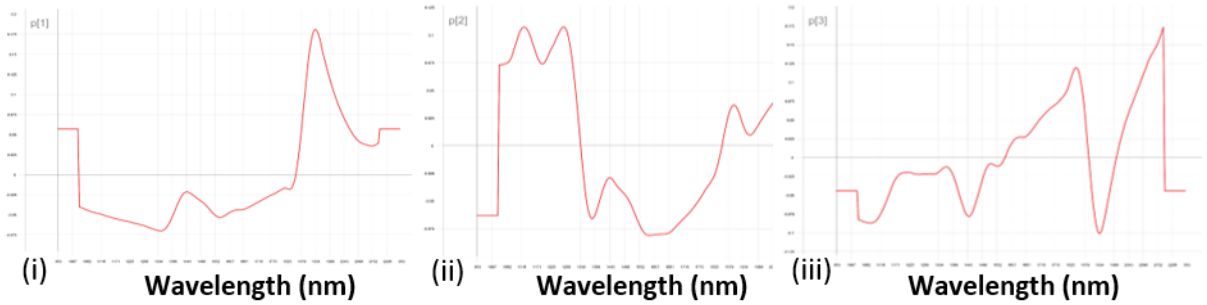

Figure B12 PCA loading lines of **Group 7**. In the positive region for (i) PC1 we observe a prominent peak at 1934 nm. In loading plot (ii) PC2, two bands are observed at 1225 and 1378 nm in the positive region. In (iii) PC3 we observe two prominent bands at 1847 and 2174 nm in the positive region. In the negative region, two bands can be seen at 1443 and 1934 nm.

Group 8: PCA results of Gram-negative bacteria – *streak*

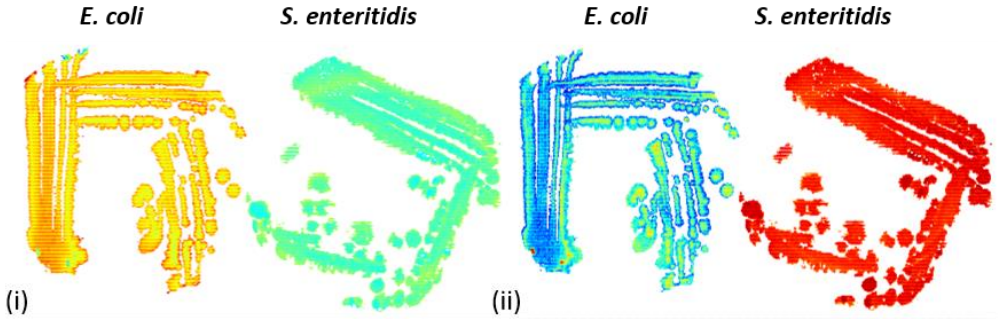

Figure B13 Score images of Gram-negative bacteria (i) before and (ii) after pre-processing is applied. SNV produced the best results for **Group 7**. The loading plots are given for (iii) PC1 with a peak observed at 1934 nm (pos. region). For (iv) the peak at 1934 nm is observed in the negative region of PC2. Lastly, for (v) PC3 a sharp peak is observed at 1934 nm (pos. region) and at 1362 and at 1443 nm (neg. region).

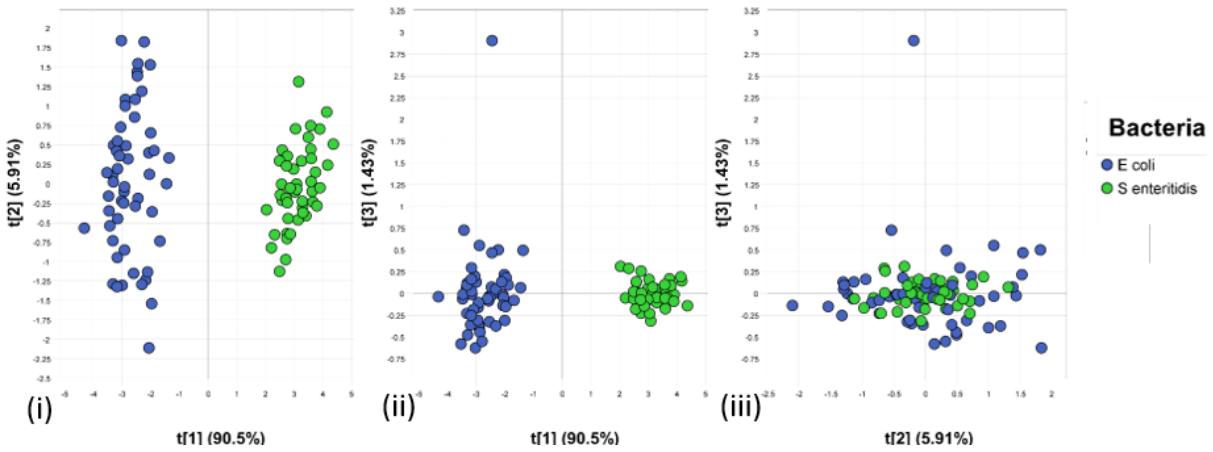

Figure B14 PCA score plots illustrating separation of the Gram-negative isolates treated with SNV. The PCA score plot (i), of PC1 (90.5%) vs. PC2 (5.91%) clear separation of the bacterial isolates. In the PCA score plot (ii), of PC1 (90.5%) vs. PC3 (1.43%), a similar observation is made. For PCA score plot (iii), PC2 (5.91%) vs. PC3 (1.43%), all the Gram-negative isolates shows overlapping for both plating techniques – no separation.

Loadings Group 8

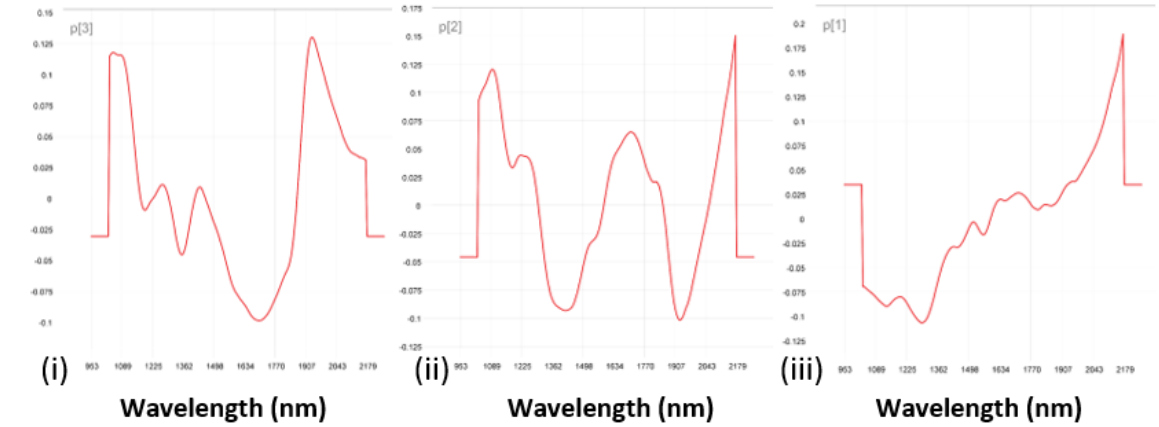

Figure B15 PCA loading lines of **Group 8**. In (i) PC1 a peak is observed in the positive region at 1934 nm. For (ii) the peak at 1934 nm observed in the positive region of PC1 is observed in the negative region in PC2. Lastly, for (iii) PC3 a peak is observed at 1934 nm (pos. region) and at 1362 and at 1443 nm (neg. region).

Loadings Group 9

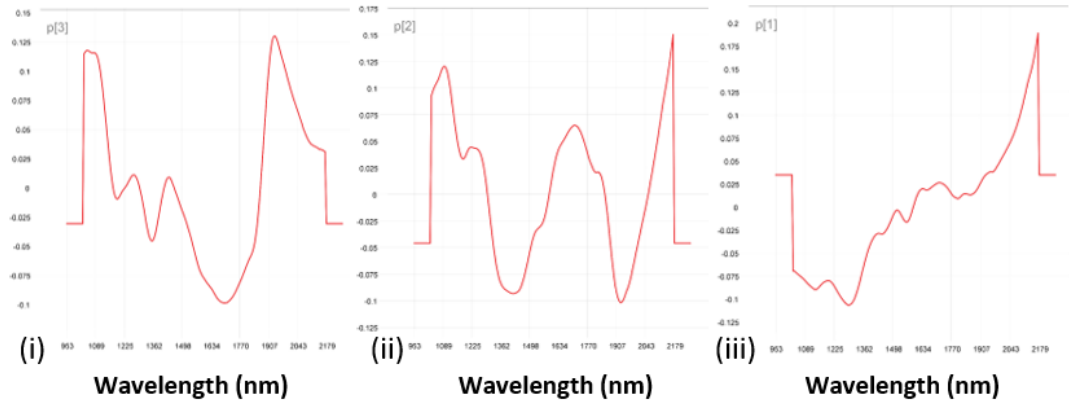

Figure B16 PCA loading lines of **Group 9**. In (i) PC1 a peak is observed in the positive region at 1928 nm. For (ii) PC2 a sharp peak is observed at 2179 nm in the positive region and in the negative region, we observe a band at 1907 nm. Lastly, for (iii) PC3 a sharp peak is located at 1934 nm (pos. region).

Group 10: PCA results of Gram-negative bacteria – *streak and spread*

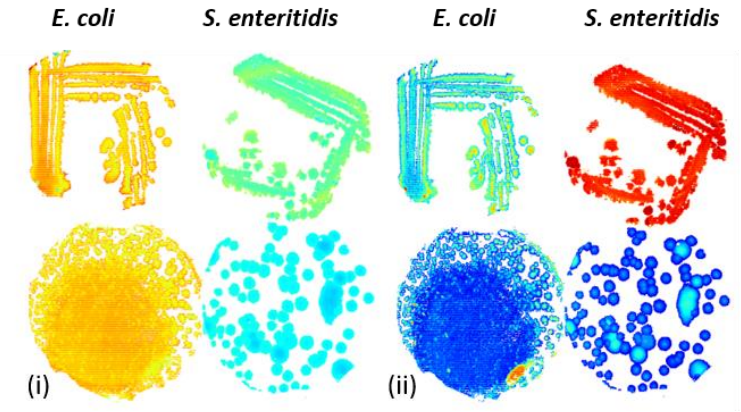

Figure B17 Score images of Gram-negative bacteria (i) before and (ii) after pre-processing is applied. SNV produced the best results for **Group 10**.

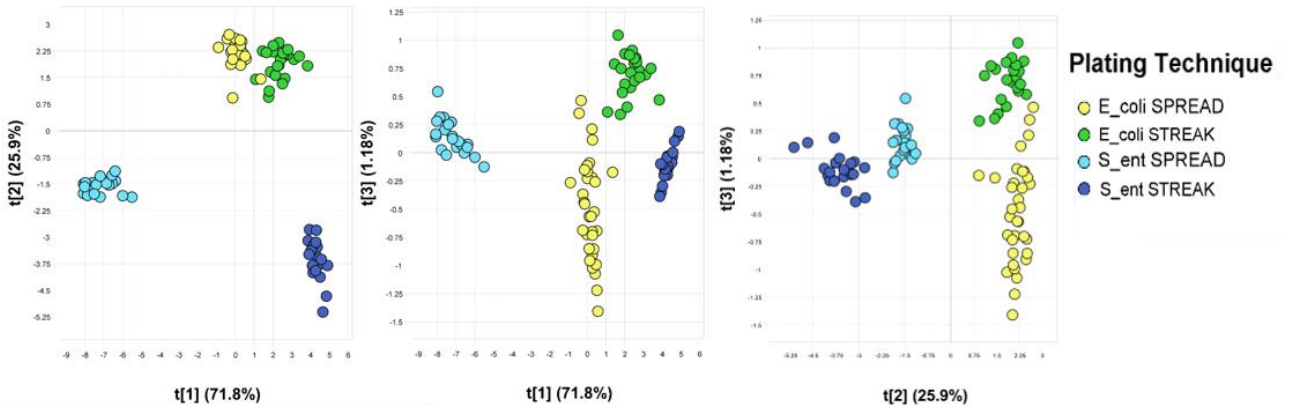

Figure B18 PCA score plots of Gram-negative isolates pre-processing - SNV. The PCA score plot (i), of PC1 (71.8%) vs. PC2 (25.9%) clear separation of the *S. enteritidis* streak and spread plates, one big cluster for the *E. coli* isolates. In PCA score plot (ii), of PC1 (71.8%) vs. PC3 (1.18%), a similar observation is made, with some spaced out data points for *E. coli*. For PCA score plot (iii), PC2 (25.9%) vs. PC3 (1.18%), the streak and spread plates of each isolate formed individual clusters.

Loadings Group 10

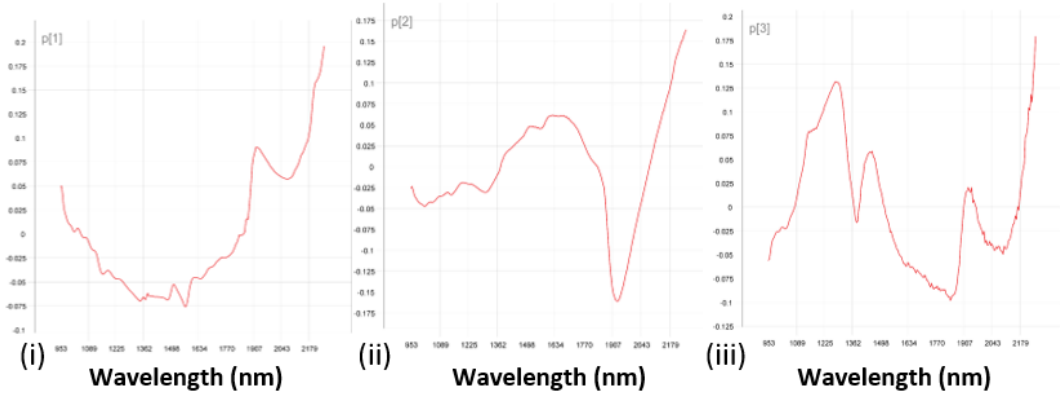

Figure B19 PCA loading lines of **Group 10**. In (i) PC1 a peak is observed in the positive region at 2174. For (ii) PC2 a sharp peak is observed at 1907 nm. Lastly, for (iii) PC3 at 1069, 1362 and 1934 nm (pos. region).

Loadings Group 11

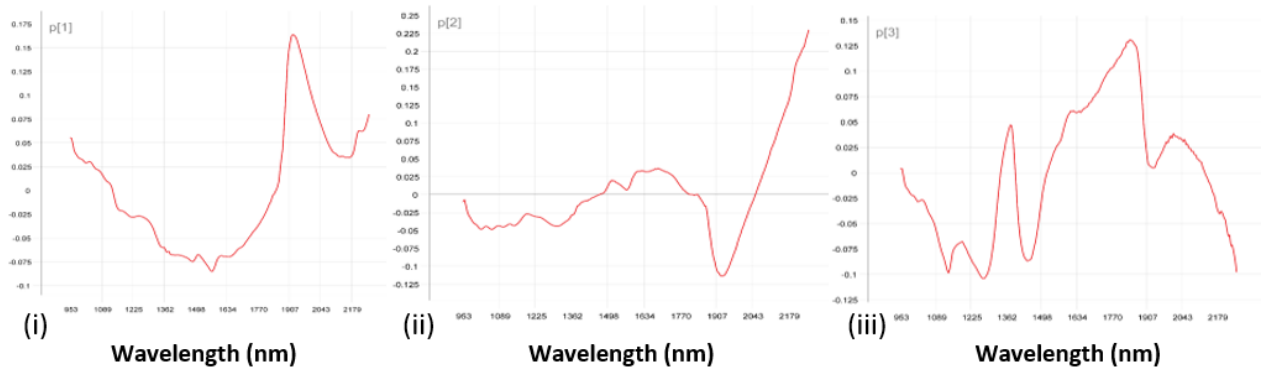

Figure B20 PCA loading lines of **Group 11**. In (i) PC1 a peak is observed in the positive region at 1928 nm and in the negative region at 2179 nm. For (ii) PC2 a sharp peak is observed at 2179 nm (inversed in PC1) and in the negative region, we observe a band at 1907 nm. Lastly, for (iii) PC3 at 1934 nm (pos. region) and at 1847 and 2179 nm (neg. region).

Digital images

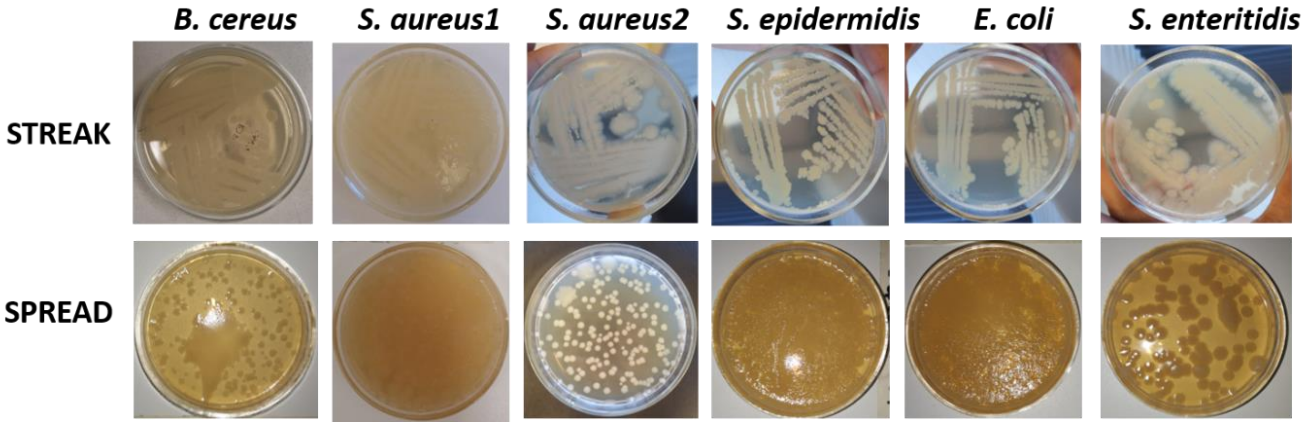

Figure B21 Digital images of foodborne pathogenic bacteria on different growth media illustrating streak and spread plate techniques used in this study. Overgrowth evident for streak plates: *S. enteritidis*; spread plates: *B. cereus, S. aureus*1, *S. epidermidis, E. coli.*

Figure B22 Accompanying RGB images of foodborne pathogenic bacteria using different plating techniques after image cleaning.